OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:51:59 ; Search time 5280.09 Seconds
(without alignments)
11342.728 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Search Notes

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1236	100.0	1239	6	BD274350	BD274350 Candida a
2	1236	100.0	2060	6	BD274349	BD274349 Candida a
3	1232.8	99.7	1239	6	AX489193	AX489193 Sequence
4	581.8	47.1	627	6	AR550053	AR550053 Sequence
c 5	438.4	35.5	110000	8	CR382135_06	Continuation (7 of
6	391	31.6	462	6	AR550055	AR550055 Sequence
c 7	228	18.4	110000	8	CR382122_03	Continuation (4 of
c 8	227.2	18.4	92211	8	CR380953_09	Continuation (10 o
9	217.4	17.6	1560	8	YSCNOFEAT	M90638 Saccharomyc
10	217.4	17.6	1739	8	YSCTFIIIA	M80611 Saccharomyc
c 11	217.4	17.6	37497	8	YSCP9677	U25841 Saccharomyc
c 12	199.2	16.1	2133	8	YSCRP026A	M33924 S.cerevisia
13	193	15.6	797	11	CNS06ING	AL400514 T7 end of
14	120.6	9.8	110000	8	AE016818_01	Continuation (2 of
c 15	101	8.2	169063	9	AC138126	AC138126 Homo sapi
c 16	101	8.2	177299	9	AC073544	AC073544 Homo sapi
17	99.4	8.0	2132	6	AX714042	AX714042 Sequence
18	99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
19	99.4	8.0	2196	6	CQ726027	CQ726027 Sequence
c 20	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
c 21	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
22	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
23	97.8	7.9	1203	9	D70831	D70831 Homo sapien
c 24	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
c 25	97.2	7.9	115995	9	AC011494	AC011494 Homo sapi
26	96.6	7.8	2320	6	AR270491	AR270491 Sequence
27	96.6	7.8	2320	9	HSU35376	U35376 Human repre
c 28	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
c 29	96.6	7.8	106811	2	CR556705	CR556705 Mus muscu
30	95	7.7	981	9	HSHZF9	X78932 H.sapiens H
c 31	94.6	7.7	186233	9	AC092329	AC092329 Homo sapi
32	94.6	7.7	189317	2	AC024483	AC024483 Homo sapi
33	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
34	93.6	7.6	1389	9	M27879	M27879 Homo sapien
c 35	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
36	93.4	7.6	199585	10	AC124426	AC124426 Mus muscu
37	92.4	7.5	3516	6	CQ726030	CQ726030 Sequence
c 38	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
39	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
40	92	7.4	2873	9	HUMHPLK	M55422 Human Kruep
41	92	7.4	2873	11	G28705	G28705 sWSS3976 Er
42	92	7.4	2873	11	HUMSWS1269	G18281 human chrom
43	92	7.4	156835	9	AC010620	AC010620 Homo sapi
c 44	92	7.4	191082	2	AC141066	AC141066 Homo sapi
c 45	92	7.4	203396	9	AC073210	AC073210 Homo sapi

ALIGNMENTS

RESULT 1
BD274350
LOCUS BD274350 1239 bp DNA linear PAT 17-JUL-2003
DEFINITION Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA protein.
ACCESSION BD274350
VERSION BD274350.1 GI:33084118
KEYWORDS JP 2002531068-A/2.
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1239)
AUTHORS Pallier,F.B., Camier,S. and Sentenac,A.
TITLE Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
JOURNAL Patent: JP 2002531068-A 2 24-SEP-2002;
AVENTIS PHARMA SA
COMMENT OS Candida albicans
PN JP 2002531068-A/2
PD 24-SEP-2002
PF 09-NOV-1999 JP 2000581204
PR 10-NOV-1998 FR 98/14147
PI FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC C07K16/14,
PC C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC G01N33/53,
PC G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
PC 00,
PC (C12N15/00,C12R1:725)
CC Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
CC protein
FH Key Location/Qualifiers
FT CDS (1). .(1236).
FEATURES Location/Qualifiers
source 1. .1239
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"
ORIGIN

Query Match 100.0%; Score 1236; DB 6; Length 1239;

```
  Best Local Similarity   100.0%;  Pred. No. 1.5e-189;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236


RESULT 2
BD274349
LOCUS       BD274349               2060 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            protein.
ACCESSION   BD274349
VERSION     BD274349.1  GI:33084117
KEYWORDS    JP 2002531068-A/1.
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 2060)
  AUTHORS   Pallier,F.B., Camier,S. and Sentenac,A.
  TITLE     Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
  JOURNAL   Patent: JP 2002531068-A 1 24-SEP-2002;
            AVENTIS PHARMA SA
COMMENT     OS   Candida albicans
            PN   JP 2002531068-A/1
            PD   24-SEP-2002
            PF   09-NOV-1999 JP 2000581204
            PR   10-NOV-1998 FR     98/14147
            PI   FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC
            C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC
            C07K16/14,
            PC   C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC
            G01N33/53,
            PC   G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
            PC   00,
            PC   (C12N15/00,C12R1:725)
            CC   Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            CC           protein
            FH   Key             Location/Qualifiers
            FT   source          1. .2060
            FT                   /organism='Candida albicans'.
FEATURES             Location/Qualifiers
     source          1. .2060
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match           100.0%;  Score 1236;  DB 6;  Length 2060;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-189;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899
```

```
Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              |||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1955
```

```
RESULT 3
AX489193
LOCUS       AX489193               1239 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6493 from Patent WO02053728.
ACCESSION   AX489193
VERSION     AX489193.1  GI:22323205
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6493 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1239
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match             99.7%;  Score 1232.8;  DB 6;  Length 1239;
  Best Local Similarity   99.8%;  Pred. No. 4.8e-189;
  Matches 1234;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236


RESULT 4
AR550053
LOCUS       AR550053                627 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5184 from patent US 6747137.
ACCESSION   AR550053
VERSION     AR550053.1  GI:53943228
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5184 08-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             47.1%;  Score 581.8;  DB 6;  Length 627;
  Best Local Similarity   97.4%;  Pred. No. 3.6e-84;
  Matches  605;  Conservative    0;  Mismatches    7;  Indels    9;  Gaps    1;

Qy        616 TGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATC 675
              ||| | || ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         13 TGTTTGGGAAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTTACCATGATC 72

Qy        676 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 132

Qy        736 GAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAA 795
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        133 GAACATTATAATATCTTCCATGATGGTAATATCCCCGATGATTTATTAAAGGAAACTGAA 192

Qy        796 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 252

Qy        856 GAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAAGATAGTCTAGATGAA 915
              |||||||||||||||||||          ||||||||||||||||||||||||||||||
Db        253 GAAACAGAGAAATTAAAAGTG---------GAAGAAGATGAAGAAGATAGTCTAGATGAA 303

Qy        916 AAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 975
              | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        304 AGAAGAAGTGATGTTAGATCAGATTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 363

Qy        976 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 423

Qy       1036 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 483

Qy       1096 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 543

Qy       1156 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 603

Qy       1216 GAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||
Db        604 GAAACATCAGTGATTTCTCGA 624


RESULT 5
CR382135_06/c
WPCOMMENT
  Sequence split into 16 fragments  LOCUS CR382135 Accession CR382135
     Fragment Name      Begin     End
     CR382135_00            1    110000
     CR382135_01       100001    210000
     CR382135_02       200001    310000
     CR382135_03       300001    410000
     CR382135_04       400001    510000
     CR382135_05       500001    610000
     CR382135_06       600001    710000
     CR382135_07       700001    810000
     CR382135_08       800001    910000
     CR382135_09       900001   1010000
     CR382135_10      1000001   1110000
     CR382135_11      1100001   1210000
     CR382135_12      1200001   1310000
     CR382135_13      1300001   1410000
     CR382135_14      1400001   1510000
     CR382135_15      1500001   1592360
Continuation (7 of 16) of CR382135 from base 600001 (CR382135 Debaryomyces hansenii chro

  Query Match             35.5%;  Score 438.4;  DB 8;  Length 110000;
  Best Local Similarity   63.4%;  Pred. No. 1.3e-61;
  Matches  724;  Conservative    0;  Mismatches  406;  Indels   12;  Gaps    3;

Qy         33 ATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACATATGAAGG 92
              ||| | ||    | || || |||||| | || || ||||| ||||||||     |   ||
Db      20471 ATCCTCAAGGGCGTCATCGTCTTCAGCTCGACCTAAAAATTATATTTGTGACGAGCCAGA 20412

Qy         93 GTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACAGTAA 152
              || |  || ||||||  |  ||||||  | ||||| ||||||  |||| | || | |||
Db      20411 CTGCGGGAAGGCCTATTCTAAACCATCGCTCTTAGAACAACATAAAAGATCACATACTAA 20352

Qy        153 TGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACA 212
              |||  |||| | ||||||  | |     |  || || ||  ||||  | ||||| |||||
```

```
Db      20351 TGAAAGACCCTTTAAATGCTCGGAAACTGGATGCGACAAGTCATTCCTAAGAAAGTCACA 20292

Qy        213 TTTGGAAACACATATTGTATC------ACATTCCGAAAAAAAACCATTCCATTGTTCAGT 266

Db      20291 TTTACAAGCACATTTGCTTTCTCATGAAGATCAAGAAAGTAAACCCTTCCAGTGTGCAAC 20232

Qy        267 GTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAA 326

Db      20231 TTGCGGCAAAGGTGTAAATACCTTACAACATTTAAAGAGACACGAAATAACACATACAAA 20172

Qy        327 GTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAG 386

Db      20171 ATCATTCATTTGTACATTTGAGGGCTGTAATGAATCATTTTATAAGCATCAGTCACTCAG 20112

Qy        387 ACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTT 446

Db      20111 ACACCATACGCTATCGGTTCATGAAAAAAAATTATTATGCAATAAGTGTAATAAATCGTT 20052

Qy        447 CACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTTA 506

Db      20051 TAACAGACCGTATCGATTGGCACAGCACAATATTAAATATCACAGTGACTCGCCGGCTTA 19992

Qy        507 TCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCA 566

Db      19991 TCAATGTGACCATCAAGGCTGTTTTGGAAATTTTATGACGTGGTCGGCTTTACAATTGCA 19932

Qy        567 TATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAA 626

Db      19931 TATAAAAACTGAGCACCCAAAGATAAAGTGTCCTATATGTGGGAAAGGTTGCGTGGGTAA 19872

Qy        627 AAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGAC 686

Db      19871 GAAAGGGCTTCGGTCACACATGAATATTCACGATGAAGAAAAGATAGTCAAATTATGGAA 19812

Qy        687 TTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTATAA 746

Db      19811 TTGTAATTATTGTAACATTGGAAAATTTGTGAAAAAAGCTGATTTGATAGATCATTACAA 19752

Qy        747 TATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAAGTGAAAAAATT 806

Db      19751 CACTTATCACGACAGAAATCTTCCTGATGATTTATTGAAACCGATCGAACGAGAACAGTT 19692

Qy        807 AGAGAACCTATTA---GATCAAGGATCGAAATTAAATAATTTGCATGAATTAGAAACAGA 863

Db      19691 AGATAAACTTCTAAGCGAGAAGGACTCTACCAATATTGATTTTAATACTTTAGAAGGATT 19632

Qy        864 GAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAAGATAGTCTAG---ATGAAAAAAG 920

Db      19631 ACAATCAAAAGGGTTTGTGGAAGTACCGTCGGATGAAGAGAAGGAAGAGAATGATTACAT 19572

Qy        921 AAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCAATAAAATCATTTACTGCTTC 980

Db      19571 GTCCAATGGTTTGCAAGCTTCACATAAATCTTTGAATTCGTTGAATTTAACTTTGGAATC 19512

Qy        981 TTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATCAATTG 1040

Db      19511 TGGAAAAGCATCTATTGTAGATCTTATTCTGAATAATTATCTGAAAAGGAAAGTTCCATG 19452

Qy       1041 TCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACATTTGAA 1100

Db      19451 TCCGAAAAAGAATTGTGATAGAATGTTCAGCCGTGATTATGATCTTGAGAGACATTTAAA 19392

Qy       1101 ATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAAGAAGA 1160

Db      19391 GTGGCATGAATCTCACTTAAAAAAGATTGAAGCGTTTTTAAAAAGTCTTGAGCTGGAAGA 19332

Qy       1161 AA 1162

Db      19331 AA 19330

RESULT 6
AR550055
```

```
LOCUS       AR550055                462 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5186 from patent US 6747137.
ACCESSION   AR550055
VERSION     AR550055.1  GI:53943230
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5186 08-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           31.6%;  Score 391;  DB 6;  Length 462;
  Best Local Similarity   98.5%;  Pred. No. 2e-53;
  Matches  405;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Db         52 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 111

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Db        112 CGTCCCAAAAAGTATATTTGTACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 171

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180

Db        172 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 231

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Db        232 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 291

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Db        292 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 351

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Db        352 AAAAGACATGAAATCACCCATACCAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 411

Qy        361 GCATTTTAT-AAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410

Db        412 ACATTTTATAAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 462


RESULT 7
CR382122_03/c
WPCOMMENT
  Sequence split into 14 fragments  LOCUS CR382122 Accession CR382122
     Fragment Name      Begin      End
     CR382122_00            1    110000
     CR382122_01       100001    210000
     CR382122_02       200001    310000
     CR382122_03       300001    410000
     CR382122_04       400001    510000
     CR382122_05       500001    610000
     CR382122_06       600001    710000
     CR382122_07       700001    810000
     CR382122_08       800001    910000
     CR382122_09       900001   1010000
     CR382122_10      1000001   1110000
     CR382122_11      1100001   1210000
     CR382122_12      1200001   1310000
     CR382122_13      1300001   1320834
Continuation (4 of 14) of CR382122 from base 300001 (CR382122 Kluyveromyces lactis strai
```

```
  Query Match             18.4%;  Score 228;  DB 8;  Length 110000;
  Best Local Similarity   58.1%;  Pred. No. 9.6e-28;
  Matches  443;  Conservative    0;  Mismatches  310;  Indels    9;  Gaps    2;

Qy         41 TATCTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATA 100
              | ||   ||| || |||   ||| | |||   ||| | |||   ||| ||||||||| | |
Db      15150 TGTCAGTTTCATCGTCACAGCGTTCTAAAGTATATTTCTGCGATTATCAAGGGTGTAACA 15091

Qy        101 AAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGAC 160
              || | |   ||||    ||  |   | | |||       ||    |    || || ||
Db      15090 AATCGTTCACAAGACCGAGTTTGCTCACGGAGCATCAGCTCAC---CGTGCATCATGGAA 15034

Qy        161 CGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAA 220
                |   | || |   |||  |||||  |||| ||||   | |||||| ||| || | |
Db      15033 TCAAACCGTTTAAATGCGACACTTGTGGCAAAGAATTTGCGAAAAAATCGCATCTGAATA 14974

Qy        221 CACATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGG 280
              |||  ||  | || ||| | ||  | ||||| |||   |||||  | ||||| ||||| |
Db      14973 GACACATGTTCTCTCATACAGATGATAAACCTTTCACATGTTCTATATGTGGGAAAGGTG 14914

Qy        281 TTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTA 340
              | |   | ||||||| || || |||||||||||| || ||||| ||||| ||   | |||
Db      14913 TCACAACCAGACAACAGTTAAAGAGACATGAAATTACACATACGAAGTCCTTCCACTGTT 14854

Qy        341 CATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTAT 400
              ||| |||    ||| | ||| |||| || ||||||| |    |  ||   ||||| || |
Db      14853 CATATGAGGGATGTAACGAATCATTCTACAAACATCCACAACTCCGATCGCATATCTTGT 14794

Qy        401 CTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAA 460
              | ||||| ||||| | | | || ||    || |||| ||    |||    | ||||
Db      14793 CCGTTCACGAAAAGAAACTTACCTGCCCTCACTGTAACAAGACGTTCCAAAGGCCTTATC 14734

Qy        461 AATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGC------TTATCAATGTG 514
              |||     | || | |   || |||||       |   || |       || ||||||||
Db      14733 GATTGAAGAATCACATAGATAAGCATCACAACCCAGAATCTACCGGAATGTACCAATGTG 14674

Qy        515 ATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAAC 574
              |   || |   |||    |    ||   ||||||||    ||||||   ||||| ||||
Db      14673 ACTTTCTTTCATGTACCGACGTATTTTCCACTTGGTCTTCTTTACAACAACATATTAAAC 14614

Qy        575 AACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTT 634
              ||   ||||| ||| |   |||||||  ||||| |||  ||||||||   ||  |||||
Db      14613 AATGCCATCCGAAATTACCATGTCCTATTTGTGGGAAACCTTGTGTTGCTGAATCAGGTT 14554

Qy        635 TATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATT 694
              ||      |||||| | | |||||| ||    |   | |  | |||  |||   ||
Db      14553 TACGGAATCATATGATGATTCATGACGAGAGTTTGGTCACTAAGAATTGGAAATGTTCTT 14494

Qy        695 ATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCC 754
              |||  |     ||||||  ||||||| | |  ||||  ||||||||  |       |
Db      14493 CATGTCCAGATACATCATTTGCGAAGAAAACTCAGCTAGTACAACATTATGAAGAAAGTC 14434

Qy        755 ATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAAG 796
              |  | | |      || ||     |||||  |   || || |
Db      14433 ACAAAGATGTAGATCCAGAATTGCTATTACTGCCGACGGACG 14392


RESULT 8
CR380953_09/c
WPCOMMENT
  Sequence split into 10 fragments  LOCUS CR380953 Accession CR380953
     Fragment Name      Begin     End
     CR380953_00            1    110000
     CR380953_01       100001    210000
     CR380953_02       200001    310000
     CR380953_03       300001    410000
     CR380953_04       400001    510000
     CR380953_05       500001    610000
     CR380953_06       600001    710000
     CR380953_07       700001    810000
     CR380953_08       800001    910000
     CR380953_09       900001    992211
Continuation (10 of 10) of CR380953 from base 900001 (CR380953 Candida glabrata strain C
```

```
  Query Match             18.4%;  Score 227.2;  DB 8;  Length 92211;
  Best Local Similarity   57.8%;  Pred. No. 1.3e-27;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    9;  Gaps    2;

Qy         44 CTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAG 103
              | |||   ||       | |  | ||| |  ||| ||||   ||| ||    ||||| || |
Db      60262 CATCTGGATCCAGTGTAAGGTCGAAGACATATCTTTGTGAATACGATAATTGTGACAAGG 60203

Qy        104 CCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGT 163
              |||  ||    | ||  ||  ||  |     ||||    |  ||        ||  | |   |
Db      60202 CCTTTACAAGGCCGTCCTTGCTCACAGAACACCAGAATAC---AGTGCATCTTGGGAGGA 60146

Qy        164 ATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACAC 223
              |    || | |    | | |||||  |   |||||    | |||     || || || | ||
Db      60145 AACCTTGGAAATGTAATCAATGTGAAAGTTCATTTACTAAAAAGATCCACTTAGAGAGAC 60086

Qy        224 ATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTA 283
              | |    |||  | ||  ||| ||| ||  |||||||| | ||||| || ||| | |
Db      60085 ACTTATACACACACACAGATGAAAGACCGTTTTATTGTTCATTTTGTGGGAAGGGGCTAA 60026

Qy        284 ATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACAT 343
              | || | |||||  ||||| |||||||  | || || || || || || || ||| ||
Db      60025 TTACTAGGCAACAACTGAAACGACATGAGGTAACTCACACCAAATCGTTCAATTGTGAAT 59966

Qy        344 TTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTG 403
              ||||  |||| | ||    || || || || |      |||||   |||||| ||| | |
Db      59965 ATGAAGGTTGTAATGAGAGTTTCTACAAGCACCCTCAATTAAGGGCTCATATTTTAGCAG 59906

Qy        404 TTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAAT 463
              |||||  | |||   ||| ||| |  |||| || |||  |||      ||||||  | |
Db      59905 TTCATTTACAAAGTCTAAAATGTCACGAATGCAACAAAAGTTTTCAAAGACCTTACAGAC 59846

Qy        464 TAGCACAACATAAATTAAAACATCAT------GGTGGATCTCCTGCTTATCAATGTGATC 517
              |   | | || | |   ||||| |||      | ||       ||||||||||||
Db      59845 TCAAAAATCACATAGCCAAACACCATAATCCAGATGTTGTCAATGCTTATCAATGTACCT 59786

Qy        518 ATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAAC 577
                || ||||| ||| | |||| |||| ||||| |  |||  ||| ||| | ||| |
Db      59785 TCAGTGTTTGTTCTAAGAGTTTCAAAACATGGTCTGCGTTAAGATTACATGTTAAAAATG 59726

Qy        578 TGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTAT 637
               ||||| ||  | ||||| |||| |||  |||||    |||||||||| || | ||||||
Db      59725 ATCATCCGAAGTTAAAATGCCCTATATGCAGTAAACCCTGTGTTGGGGAAGACGGTTTAA 59666

Qy        638 CTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATT 697
              |   || |||  ||  || |||||      |    | ||  ||||  ||| ||   |
Db      59665 ATATGCACATGAAAATCCACGATGAAAACCTTGTTTCAAGAAATTGGAAGTGTCATATAT 59606

Qy        698 GTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCCATG 757
              || |||     |  |||||| ||||||    ||| || | || |||||  |  | |  |||
Db      59605 GTAATGATCAGTCATTTGCTAAGAAATTAGAACTATTGGATCATTACAGCAACAGTCATA 59546

Qy        758 ATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAAGTGAAAAAATTAGAGA 811
              ||   | |||||  |    || |||||  |  |  |  ||||     |  | || |
Db      59545 GTGAAGAAATCCCGGCCTATCTATTAGAACAGAAGGTAGTGCCTTATGTTGAAA 59492


RESULT 9
YSCNOFEAT
LOCUS       YSCNOFEAT               1560 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  Saccharomyces cerevisiae transcription factor IIIA and RNA
            polymerase subunit RPB6 genes.
ACCESSION   M90638
VERSION     M90638.1  GI:172046
KEYWORDS    RNA polymerase; transcription factor IIIA.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Woychik,N.A. and Young,R.A.
  TITLE     Genes encoding transcription factor IIIA and the RNA polymerase
```

common subunit RPB6 are divergently transcribed in Saccharomyces cerevisiae
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)
MEDLINE 92237295
PUBMED 1570325
COMMENT Original source text: Saccharomyces cerevisiae (strain S288C) DNA.
FEATURES Location/Qualifiers
source 1. .1560
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="S288C"
/db_xref="taxon:4932"
ORIGIN chromosome 14.

Query Match 17.6%; Score 217.4; DB 8; Length 1560;
Best Local Similarity 59.4%; Pred. No. 1.3e-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

```
Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
Db        368 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 427

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
Db        428 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 484

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
Db        485 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 544

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
Db        545 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 604

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
Db        605 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 664

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
Db        665 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 724

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
Db        725 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 784

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
Db        785 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 844

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
Db        845 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 904

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
Db        905 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 964

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
Db        965 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1024

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
Db       1025 CTAGAAAACATGATCTTCTCACGCATTAT 1053
```

RESULT 10
YSCTFIIIA
LOCUS YSCTFIIIA 1739 bp DNA linear PLN 13-SEP-1996
DEFINITION Saccharomyces cerevisiae transcription factor IIIA (TFIIIA) gene, complete cds.
ACCESSION M80611
VERSION M80611.1 GI:172902
KEYWORDS transcription factor IIIA; zinc-finger protein; zinc-finger transcription factor.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1739)
AUTHORS Archambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D. and Segall,J.
TITLE The deduced sequence of the transcription factor TFIIIA from Saccharomyces cerevisiae reveals extensive divergence from Xenopus TFIIIA
JOURNAL J. Biol. Chem. 267 (5), 3282-3288 (1992)
MEDLINE 92147684
PUBMED 1737784
FEATURES Location/Qualifiers
source 1. .1739
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="W303-1A"
/db_xref="taxon:4932"
/chromosome="XVI"
gene 234. .1523
/gene="TFIIIA"
CDS 234. .1523
/gene="TFIIIA"
/note="zinc finger containing protein"
/codon_start=1
/product="transcription factor IIIA"
/protein_id="AAB08014.1"
/db_xref="GI:172903"
/translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
ORIGIN

Query Match 17.6%; Score 217.4; DB 8; Length 1739;
Best Local Similarity 59.4%; Pred. No. 1.3e-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

```
Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
Db        367 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 426

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
Db        427 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 483

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
Db        484 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 543

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
Db        544 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 603

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
Db        604 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 663

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
Db        664 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 723

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
Db        724 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 783
```

```
Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
              |||||||      ||  | |     | |||| ||||||  |  | |||||||||  |||
Db        784 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 843

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
              | || |  | ||||||  ||| ||||  |||||||||  |    ||||| ||| | ||||
Db        844 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 903

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
              ||||||  ||| | |||   ||||| ||| |||| ||||||     || ||| | | ||
Db        904 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 963

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
              |||| || ||     | | |||||  ||||  ||| ||   |||  ||    |  |||
Db        964 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1023

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
              | |  ||| ||||  |  |    ||||||
Db       1024 CTAGAAAACATGATCTTCTCACGCATTAT 1052


RESULT 11
YSCP9677/c
LOCUS       YSCP9677              37497 bp    DNA     linear   PLN 01-AUG-1997
DEFINITION  Saccharomyces cerevisiae chromosome XVI cosmid 9677.
ACCESSION   U25841 U00094
VERSION     U25841.1  GI:786295
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 37497)
  AUTHORS   Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Du,Z.,
            Favello,A., Fulton,L., Gattung,S., Greco,T., Kirsten,J., Kucaba,T.,
            Hallsworth,K., Hawkins,J., Hillier,L., Jier,M., Johnson,D.,
            Johnston,L., Langston,Y., Latreille,P., Le,T., Mardis,E.,
            Menezes,S., Miller,N., Nhan,M., Pauley,A., Peluso,D., Rifken,L.,
            Riles,L., Taich,A., Trevaskis,E., Vignati,D., Wilcox,L.,
            Wohldman,P., Vaudin,M., Wilson,R. and Waterston,R.
  TITLE     The sequence of a portion of the right arm of Saccharomyces
            cerevisiae chromosome XVI
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 37497)
  AUTHORS   Miller,N.
  TITLE     The sequence of S. cerevisiae cosmid 9677
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 37497)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1995) Robert Waterston
COMMENT     Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA
                      e-mail: mj@sequencer.wustl.edu

            NEIGHBORING COSMID INFORMATION:
            This sequence includes nucleotides 1-37497 of cosmid 9677.  The
            cosmid on the left is 9705.
FEATURES             Location/Qualifiers
     source          1. .37497
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C (AB972)"
                     /db_xref="taxon:4932"
                     /map="XVI"
                     /note="baker's yeast"
     gene            complement(2366. .3655)
                     /gene="TFC2"
     CDS             complement(2366. .3655)
                     /gene="TFC2"
                     /note="P9677.9"
                     /codon_start=1
                     /product="Transcription factor TFIIIA (PIR accession
                     number S20050)"
                     /protein_id="AAB64615.1"
                     /db_xref="GI:786305"
                     /translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
                     RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
                     SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
                     HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
                     HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
                     HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
                     EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
                     ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
     gene            3890. .4433
                     /gene="RPB6"
     CDS             join(3890. .3909,3986. .4433)
                     /gene="RPB6"
                     /note="P9677.8"
                     /codon_start=1
                     /product="23 kD polypeptide of RNA Polymerases I, II, and
                     III (Swiss Prot. accession number P20435)"
                     /protein_id="AAB64616.1"
                     /db_xref="GI:786304"
                     /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                     GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                     MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
     gene            complement(4621. .5112)
                     /gene="P9677.10"
     CDS             complement(4621. .5112)
                     /gene="P9677.10"
                     /note="Similar to calmodulins (e.g., Swiss Prot. accession
                     number P13868) and calmodulin-related proteins (e.g.,
                     Swiss Prot. accession number P27482) from several
                     organism."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB64617.1"
                     /db_xref="GI:786306"
                     /translation="MDHSESLTFNQLTQDYINKLKDAFQMLDEDEDGLISRGDLTKIY
                     ATLGKTLTDEEWSKMVPDNDTSTAEVGEEGVSFPIFLSIMGKNLSQFPEREELEESLK
                     AIGRGHDLNVPLNEVIDSLKEAGFENPEEEFAKLFKLFTTNQQATEERTFRGKLFLDS
                     ITD"
     gene            5297. .9595
                     /gene="SKI3"
     CDS             5297. .9595
                     /gene="SKI3"
                     /note="P9677.7"
                     /codon_start=1
                     /product="Contains 8 copies of the TPR domain"
                     /protein_id="AAB64618.1"
                     /db_xref="GI:786303"
                     /translation="MSDIKQLLKEAKQELTNRDYEETIEISEKVLKLDPDNYFAHIFL
                     GKALSSLPASNNVSSNRNLERATNHYVSAAKLVPDNLLAWKGLFLLFRTTEVVPDILS
                     YDEYFDLCGQYADALLKQEQSQVELINDIKLLKKTHPDCQKAFYQHLKPGSLMAETIG
                     RHLSTPQDALLNLIKILSNIETTEIGKTLSQNRLKLKASDPDYQIKLNSFSWEIIKNS
                     EIDQLYNQLVNILADDQKRSEIENQWLEYRIKVLKSMPLDVKKDFFTKVKEMVEDMVL
                     VNHQSLLAWQKYFEWTDYEDLDNMDAPLIIKYFKKFPKDPLAMILYSWLSSKLSKYDI
                     KSLESANKPPEGHKKTEKETDIKDVDETNEDEVKDRVEDEVKDRVEDEVKDQDEEAKE
                     DEEEDLDDIEIGLLEEEVVTVLTENIVKCKNNILAHRILCQYYLLTKEYEAALPYIKN
                     GISLIAYNIKDLGVHLPLTKREFSLDLATVYTYVDAPKDHNAALKLYDNILSGDFSNI
                     QAKMGKGIIFIERKNWKDAMTLLTQVHEQSPNNLEVLSELSWSKAHMGYMDEALAGLD
                     TVIKGIKGMDLRSIDFRALNLWRQAKVYIMKHASINDAKQENVKCAFKLLIQSIKILD
                     TFAPGFSTLGDIYCHYYKDHLRAFKCYFKAFDLDAGDYTAAKYITETYASKPNWQAAS
                     SIASRLIKGEKAKAELRSNNWPFRVVGIAHLEKQEESDSIEWFQSALRVDPNDVESWV
                     GLGQAYHACGRIEASIKVFDKAIQLRPSHTFAQYFKAISLCDVGEYLESLDILEKVCQ
                     EAATEESFQIGLVEVLMRCSLDLYSQGFLLKSVSIAKDTIERIKIIISELKCENQQVW
                     IYLSQVLRLFIWIESKVDTLPVESLVSIFENSQFSGSEEIDSVDNIKIDTLLDSTTDD
                     NVSIACKFLILASKYSVSDQKFTDIAGTVRASYWYNIGISELTAFITLKEPQYRDAAI
                     FAFKKSIQLQSNTSETWIGLGIATMDINFRVSQHCFIKATALEPKATNTWFNLAMLGL
                     KKKDTEFAQQVLNKLQSLAPQDSSPWLGMALILEEQGDIIGSSKLFAHSFILSNGRSK
                     AAQFMYAKNVLENHINNGDDERDIETVEKLTTASIALEQFFKKSPDSQFALQCALLTL
                     ERLHHYENANELANRLIGILEKKFEKTQDERELFNFAIIKGQFARIHLGLGNFELSIE
```

```
                NADLSQGIISESSDEKSMKTKISNHICLGLSYFFLNDFDQTLNQFQELLSISKDSKHL
                VVLIAKVLYDVGESDTKEIALQELTEYIATSGADLLVTLTIAAMSILDDKREDLSIIL
                EELKALPLSKQIIDKHKDAPYLIEEITKRLYRNDTGKQVWQRSAYFFPNNLKVWERLD
                KNIQRRIASNGQNKVTAEEMSKLYCESKNLRSIQRGMFLCPWNVTAVKALNECF"
gene            complement(9710. .11674)
                /gene="RPC3"
CDS             complement(9710. .11674)
                /gene="RPC3"
                /note="P9677.11"
                /codon_start=1
                /product="74 kD subunit of RNA Polymerase III (Swiss Prot.
                accession number P32349)"
                /protein_id="AAB64619.1"
                /db_xref="GI:786307"
                /translation="MDELLGEALSAENQTGESTVESEKLVTPEDVMTISSLEQRTLNP
                DLFLYKELVKAHLGERAASVIGMLVALGRLSVRELVEKIDGMDVDSVKTTLVSLTQLR
                CVKYLQETAISGKKTTYYYYNEEGIHILLYSGLIIDEIITQMRVNDEEEHKQLVAEIV
                QNVISLGSLTVEDYLSSVTSDSMKYTISSLFVQLCEMGYLIQISKLHYTPIEDLWQFL
                YEKHYKNIPRNSPLSDLKKRSQAKMNAKTDFAKIINKPNELSQILTVDPKTSLRIVKP
                TVSLTINLDRFMKGRRSKQLINLAKTRVGSVTAQVYKIALRLTEQKSPKIRDPLTQTG
                LLQDLEEAKSFQDEAELVEEKTPGLTFNAIDLARHLPAELDLRGSLLSRKPSDNKKRS
                GSNAAASLPSKKLKTEDGFVIPALPAAVSKSLQESGDTQEEDEEEEDLDADTEDPHSA
                SLINSHLKILASSNFPFLNETKPGVYYVPYSKLMPVLKSSVYEYVIASTLGPSAMRLS
                RCIRDNKLVSEKIINSTALMKEKDIRSTLASLIRYNSVEIQEVPRTADRSASRAVFLF
                RCKETHSYNFMRQNLEWNMANLLFKKEKLKQENSTLLKKANRDDVKGRENELLLPSEL
                NQLKMVNERELNVFARLSRLLSLWEVFQMA"
gene            12014. .13120
                /gene="QCR2"
CDS             12014. .13120
                /gene="QCR2"
                /note="P9677.6"
                /codon_start=1
                /product="Ubiquinol-cytochrome C reductase core protein 2"
                /protein_id="AAB64620.1"
                /db_xref="GI:786302"
                /translation="MLSAARLQFAQGSVRRLTVSARDAPTKISTLAVKVHGGSRYATK
                DGVAHLLNRFNFQNTNTRSALKLVRESELLGGTFKSTLDREYITLKATFLKDDLPYYV
                NALADVLYKTAFKPHELTESVLPAARYDYAVAEQCPVKSAEDQLYAITFRKGLGNPLL
                YDGVERVSLQDIKDFADKVYTKENLEVSGENVVEADLKRFVDESLLSTLPAGKSLVSK
                SEPKFFLGEENRVRFIGDSVAAIGIPVNKASLAQYEVLANYLTSALSELSGLISSAKL
                DKFTDGGLFTLFVRDQDSAVVSSNIKKIVADLKKGKDLSPAINYTKLKNAVQNESVSS
                PIELNFDAVKDFKLGKFNYVAVGDVSNLPYLDEL"
gene            14493. .15410
                /gene="P9677.5"
CDS             14493. .15410
                /gene="P9677.5"
                /note="Belongs to the MIP family of transmembrane channel
                proteins."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB64621.1"
                /db_xref="GI:786301"
                /translation="MSSNDSNDTDKQHTRLDPTGVDDAYIPPEQPETKHHRFKISRDT
                LRDHFIAAVGEFCGTFMFLWCAYVICNVANHDVALVAAPDGSHPGQLIMIAIGFGFSV
                MFSIWCFAGVSGGALNPAMSLSLCLARAVSPTRCVVMWVSQIVAGMAAGGAASAMTPG
                EVLFANSLGLGCSRTRGLFLEMFGTAILCLTVLMTAVEKRETNFMAALPIGISLFIAH
                VALTAYTGTGVNPARSLGAAVAARYFPHYHWIYWIGTLLGSILAWSVWQLLQILDYTT
                YVTAEKAASTKEKAQKKGETSSSSAVAEV"
gene            complement(15542. .16012)
                /gene="P9677.12"
CDS             complement(15542. .16012)
                /gene="P9677.12"
                /note="Similar in C-terminus to Gentamicin
                3'-acetyltransferase of P. aeruginosa (Swiss Prot.
                accession number P23181)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB64622.1"
                /db_xref="GI:786308"
                /translation="MSNTSEDNITVRFVTENDKEGWQRLWKSYQDFYEVSFPDDLDDF
                NFGRFLDPNIKMWAAVAVESSSEKIIGMINFFNHMTTWDFKDKIYINDLYVDENSRVK
                GAGGKLIQFVYDEADKLGTPSVYWCTDESNHRAQLLYVKVGYKAPKILYKRKGY"
gene            complement(16937. .19570)
                /gene="P9677.13"
CDS             complement(16937. .19570)
                /gene="P9677.13"
                /note="Similar to S. cerevisiae hypothetical protein
                HRD799 (PIR accession number S45161) and S. pombe
```

```
 Query Match             17.6%;  Score 217.4;  DB 8;  Length 37497;
 Best Local Similarity   59.4%;  Pred. No. 6.3e-26;
 Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
               | || || |  ||| | ||    ||||| || |||||||| || | ||   |||| |||
Db       3522 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 3463

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
               | ||     ||||   |  || |     | | |  |        || |      ||
Db       3462 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 3406

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
               | ||||  ||| | ||  | | |||    || || || | |||  |      | ||||| |
Db       3405 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 3346

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
               | |  |||||||||||| |||||    ||||| ||||| || |   |||| || ||  |||
Db       3345 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 3286

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
               |  |||| ||| | || ||||| || || || |  ||| ||   |||   ||  |
Db       3285 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 3226

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
               ||| || || |||| |   |||||    ||||| ||||||||||||  | | | | |||
Db       3225 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 3166

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
               | |||  ||| || ||||||   ||     | || |  |   ||  | | || |  |  |
Db       3165 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 3106

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
               |||||||      ||  | |      | |||| ||||||  |  | |||||||||  |||
Db       3105 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 3046

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
               | ||  |  |  ||||||  ||| ||||  |||||||||  |    |||||  ||| | ||||
Db       3045 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 2986

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
               ||||||  ||| | |||   ||||| ||| |||| ||||||      || ||| | | ||
Db       2985 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 2926

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
               |||| || ||      | | |||||  |||| ||| ||   |||  ||     |  |||
Db       2925 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 2866

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
               | |  ||| ||||  |  |    ||||||
Db       2865 CTAGAAAACATGATCTTCTCACGCATTAT 2837


RESULT 12
YSCRP026A/c
LOCUS       YSCRP026A              2133 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  S.cerevisiae RNA polymerase II sixth subunit (RP026) gene, complete
            cds.
ACCESSION   M33924
VERSION     M33924.1  GI:172452
KEYWORDS    RNA polymerase II; c-myc proto-oncogene; transmembrane protein;
            tyrosine kinase.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2133)
```

```
  AUTHORS   Archambault,J., Schappert,K.T. and Friesen,J.D.
  TITLE     A suppressor of an RNA polymerase II mutation of Saccharomyces
            cerevisiae encodes a subunit common to RNA polymerases I, II, and
            III
  JOURNAL   Mol. Cell. Biol. 10 (12), 6123-6131 (1990)
  MEDLINE   91061718
   PUBMED   2247052
COMMENT     Original source text: S.cerevisiae single-copy plasmid DNA, clone
            2.
            Draft entry and computer-readable sequence for [Unpublished (1990)]
            kindly submitted
            by J.D.Friesen, 01-MAY-1990.
               The Hospital for Sick Children
               555 University Avenue
               Toronto, Ontario, CANADA M5G 1X8
               e-mail: jim@sickkids.toronto.edu.
FEATURES             Location/Qualifiers
     source          1. .2133
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     CDS             join(1026. .1045,1122. .1569)
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /codon_start=1
                     /protein_id="AAA34989.1"
                     /db_xref="GI:172453"
                     /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                     GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                     MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
     exon            <1026. .1045
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=1
     intron          1046. .1121
                     /note="RP026 intron A"
     exon            1122. .>1569
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=2
ORIGIN      Chromosome XVI.

  Query Match             16.1%;  Score 199.2;  DB 8;  Length 2133;
  Best Local Similarity   61.4%;  Pred. No. 1.1e-22;
  Matches  340;  Conservative    0;  Mismatches  208;  Indels    6;  Gaps    1;

Qy        171 TACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGT 230
              | |||||     | ||||  ||| | ||  | | |||     || || || | |||  |
Db        556 TTCAGTGTGATAGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTA 497

Qy        231 ATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCG 290
              | ||||| || |  |||||||||| |||||    ||||| ||||| || |   ||||
Db        496 TACGCATTCTGATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCG 437

Qy        291 ACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAA 350
              || ||  ||||  |||| ||| | || ||||| || || || | ||| ||   |||
Db        436 CCAGCAACTGAAGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGG 377

Qy        351 TTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410
              ||  |      ||| || || |||| |   |||||    ||||| ||||||||||||
Db        376 ATGCAACCTCCGATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTT 317

Qy        411 AAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACA 470
              | | | | |||| |||  ||| || ||||||   ||     | || |  |   ||  | |
Db        316 ACATAAACTAACCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAA 257

Qy        471 ACATAAATTAAAACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGG 524
              || |  | ||||||||      ||  | |     | |||| ||||||  |  | ||||
Db        256 TCACATTTCTAAACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGG 197

Qy        525 TTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCC 584
              |||||  ||| | || |  | ||||||  ||| |||| |||||||| |    |||||
Db        196 TTGTTGCAAAGAGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCC 137

Qy        585 AAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACA 644
              ||| | ||||||||||  ||| | |||   ||||| ||| |||| ||||||      ||

Db        136 TAAATTAAAATGTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCA 77

Qy        645 TATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGT 704
              ||| | | |||||| || ||     | | |||||  ||||  ||| ||   |||  ||
Db         76 CATGATTATTCATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGA 17

Qy        705 GGGGAAATTTGCAA 718
                |   ||| | |
Db         16 TATGTCTTTTTCTA 3


RESULT 13
CNS06ING
LOCUS       CNS06ING                 797 bp    DNA     linear   STS 30-NOV-2001
DEFINITION  T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL400514
VERSION     AL400514.1  GI:12156635
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 797)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
  MEDLINE   20584715
   PUBMED   11152880
REFERENCE   3  (bases 1 to 797)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .797
                     /organism="Saccharomyces bayanus"
                     /mol_type="genomic DNA"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA022B07"
                     /clone_lib="AS0AA"
                     /note="end : T7"
     misc_feature    <2. .>796
                     /note="similar to Saccharomyces cerevisiae ORF YPR186c [
                     PZF1 ; TFIIIA (transcription initiation factor) ]"
                     /evidence=not_experimental
```

```
ORIGIN

  Query Match            15.6%;  Score 193;  DB 11;  Length 797;
  Best Local Similarity  59.3%;  Pred. No. 1.3e-21;
  Matches  347;  Conservative    1;  Mismatches  231;  Indels    6;  Gaps    1;

Qy        178 GACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACAT 237
              ||| ||||||  ||| | ||| | | :||    ||| | || |  ||  |  | || |||
Db         83 GACAATTGTGCAAAATCCTTTGTTAARAAGAGTCATCTAGAGAGGCACTTGTTTTCGCAT 142

Qy        238 TCCGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACAC 297
              || || |  || ||||||| |||||    |||||| || ||||| |   ||||||| ||
Db        143 TCTGATACGAAGCCATTCCAATGTTCTTATTGTGGCAAGGGGGTGACAACTCGACAGCAA 202

Qy        298 TTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAA 357
              |||||  | || ||| | || ||||| || || |||   ||| |    ||    ||| |
Db        203 TTGAAGCGCCACGAAGTTACACATACCAAATCTTTTGTTTGTCCGGAAGATGGATGTGAT 262

Qy        358 GAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACA 417
              |||||||||| || |     |||||  ||||||  ||||||| ||   | | | |
Db        263 CTCCGATTTTATAAGCACCCTCAATTAAGGGCTCATATTCTATCTGTCCACCTAGACAAA 322

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              ||||| |||  |||||||| ||||  |||     |||| |  |   ||    | ||||
Db        323 TTAACCTGTCCACAATGTAGTAAAARTTTTCAGAGACCCTACAGGCTAAGGAATCATATC 382

Qy        478 TTAAAACATCATGGT------GGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTT 531
              |  |||||||||| |      |       | | |||||||||  |    |||| |||| |
Db        383 TCTAAACATCATGATCCTGAAGTGGTAAATCCGTATCAATGTACTTTCGCTGGCTGTTGT 442

Qy        532 AAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTT 591
              | | | ||   |  ||||||   || ||||  ||||| ||| |     ||||| ||| |
Db        443 ACAGAGTACCGTATATGGTCACAGTTGCAATCACATATCAAAAATGACCATCCTAAATTG 502

Qy        592 AAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTA 651
              ||||| || |  ||    ||  |||||| ||| |||| ||| ||     ||||||  |
Db        503 AAATGCCCCATTTGCAACAAGCCTTGTGTAGGGGAAAATGGTCTACAAATGCATATGATC 562

Qy        652 AGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAA 711
              | |||||||||  |||    | |  || |  ||||  ||| ||  ||||  |
Db        563 ATTCATGATGAATCTTTGGTAACTAAGAACTGGAAATGTCATATTTGTTCTAATACATCT 622

Qy        712 TTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCCAT 756
              ||    |  ||  | ||  |  ||   || ||  | ||    |||
Db        623 TTCTTGAGAAAGCACGACCTCCTTACTCACTACCAGATGACGCAT 667


RESULT 14
AE016818_01
WPCOMMENT
  Sequence split into 16 fragments  LOCUS AE016818 Accession AE016818
     Fragment Name      Begin     End
     AE016818_00            1    110000
     AE016818_01       100001    210000
     AE016818_02       200001    310000
     AE016818_03       300001    410000
     AE016818_04       400001    510000
     AE016818_05       500001    610000
     AE016818_06       600001    710000
     AE016818_07       700001    810000
     AE016818_08       800001    910000
     AE016818_09       900001   1010000
     AE016818_10      1000001   1110000
     AE016818_11      1100001   1210000
     AE016818_12      1200001   1310000
     AE016818_13      1300001   1410000
     AE016818_14      1400001   1510000
     AE016818_15      1500001   1519138
Continuation (2 of 16) of AE016818 from base 100001 (AE016818 Eremothecium gossypii ATCC

  Query Match             9.8%;  Score 120.6;  DB 8;  Length 110000;
  Best Local Similarity  50.1%;  Pred. No. 1.9e-10;
  Matches  386;  Conservative    0;  Mismatches  374;  Indels   11;  Gaps    3;

Qy         25 TCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACA 84
              ||| | ||   | |     ||  | | ||      | || | ||||| | ||   |||
Db      17349 TCGTTGTCGAGTGTTGAGTCCACGGCGTCGAGCACGCGGCGCAAAACGCATTACTGCGAC 17408

Qy         85 TATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACA-TTTAAGAAC 143
              || || || ||  | || || |  |  || || ||  |  |   | | ||     |||
Db      17409 TACGAGGGCTGCTACAAGGCGTTCACGCGGCCGTCCCTGCTGACGGAGCACCAGCAGACG 17468

Qy        144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAG 203
              | ||  |  |     | | |||   || |    ||  | || |      |  |||  ||
Db      17469 GCGCACCAGGGCATCAGGGCATACCAGTGC----GAGCAGTGCGGGCGCGGGTTTACCAA 17524

Qy        204 AAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTC 263
              || || || |||||     ||  |  | || || || || | ||| || |||   ||  |
Db      17525 GAAGTCGCACTTGGAGCGGCACCTGTTCTCGCACTCTGAGACAAAGCCGTTCAGCTGCAC 17584

Qy        264 AGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATAC 323
              || ||||| || ||||| |   | || || ||  |||   | || || ||||| || ||
Db      17585 GGTATGTGGCAAGGGGGTGACCACGCGGCAGCAACTGAGGCGGCACGAGATCACACACAC 17644

Qy        324 AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTT 383
              |||||| || || ||  |     ||     ||     ||||| || || || || |     ||
Db      17645 AAAGTCGTTCAAGTGCCCGCACGAGGGCTGCGGCGAAGCGTTCTACAAGCACCCGCAGTT 17704

Qy        384 AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGT 443
              |     ||  |  |  | || || ||  |||   | ||||| | ||| ||| | ||
Db      17705 GCGTTCGCACGTGCTGGCGGTGCACGAGCAAAAGCTGACGTGCACACACTGTGACAAGCG 17764

Qy        444 TTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCC--- 500
              |||   || || |     ||   | ||| |     || ||||| ||     |  |
Db      17765 GTTCCAACGGCCATACCGTTTGAAAACACACATCGCCAAGCATCACGGGCCTGCATCGCA 17824

Qy        501 ---TGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATT 557
                 || |||||      |  | || ||  |  |     ||  | || ||||| |   |
Db      17825 GTTCCGGTACCAATGCACGAACGCCGGCTGCGTACAGTGCTTTGAGACGTGGTCTGCGCT 17884

Qy        558 ACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTG 617
              ||    ||  |  |      || || || ||   |||        || || ||    ||
Db      17885 GCAGCAGCACCTTCACACGGACCACCCGAAGCTACCATGCGGGGTGTGCGGGAAGCTGTG 17944

Qy        618 TGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAA 677
              || ||| | |  ||  |       || |||      || || || ||     |||| ||
Db      17945 CGTGGGGGAGACGGGACTGCAGATGCACATGACTGTGCACGACGAGTCGCGGGTGATTAA 18004

Qy        678 AATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGA 737
              ||  |||| ||   |  |||   |      | |  || ||||     ||  |  |||
Db      18005 AAACTGGAAGTGCTCTGTGTGTAGCGACACCACCTACGCGAAGATGGCCGATCTGCTTGC 18064

Qy        738 ACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGA 788
              |||||||  |    ||| | || |  |||||  | || ||  || || |
Db      18065 GCATTATATGAAAACCCACGGTGATTCGATCCCCAAGGAATTGATAGAGCA 18115


RESULT 15
AC138126/c
LOCUS       AC138126               169063 bp    DNA     linear   PRI 17-DEC-2002
DEFINITION  Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
ACCESSION   AC138126
VERSION     AC138126.1  GI:27151357
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169063)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169063)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
```

```
            Drive, Walnut Creek, CA 94598, USA
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .169063
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-274A19"
ORIGIN

  Query Match             8.2%;  Score 101;  DB 9;  Length 169063;
  Best Local Similarity   50.8%;  Pred. No. 2.4e-07;
  Matches  298;  Conservative    0;  Mismatches  280;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |           || ||  |  |  || |   |      |   | | |
Db      80134 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 80075

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | || |    | || || |||||  ||||  ||||| |   ||||| | |||  | |
Db      80074 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 80015

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||   | ||  || | |   ||   ||| || | ||||      || || |||
Db      80014 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 79961

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |       ||||| | |  |||||||  |||   ||| | | | |
Db      79960 GGCAAAGCCTTTAACCATTCCTCACATCTTGCTACACATAAGGTAATTCATACTGGAGAG 79901

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  ||||||||||   ||||        |||| | |  |||
Db      79900 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 79841

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  | |||||              |    || |||||| |||    |||| |||
Db      79840 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 79781

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |       |    | |   ||||| | | | |   |  |  ||| |||   |  |
Db      79780 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 79721

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  ||||||  |  | |||          |||||| ||  |  ||||||||  |||
Db      79720 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 79661

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||      ||    |       ||   |         ||||||| ||  ||
Db      79660 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 79601

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              ||||| |       ||  ||     |||  | ||    | | | || |
Db      79600 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTACAAGT 79554


Search completed: February  9, 2005, 05:36:20
Job time : 5288.09 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:43:04 ; Search time 640.722 Seconds
(without alignments)
11419.610 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1236	100.0	2060	3	AAA15398	Aaa15398 DNA encod
	2	1232.8	99.7	1239	6	ABZ32206	Abz32206 Candida a
	3	99.4	8.0	2132	10	ADA53158	Ada53158 Human cod
	4	96.8	7.8	1705	9	ACA98970	Aca98970 cDNA enco
	5	96.6	7.8	2320	10	ACA56456	Aca56456 Human sig
	6	96.6	7.8	2320	12	ADI56252	Adi56252 Human pol
	7	95.2	7.7	4563	5	AAS91317	Aas91317 DNA encod
	8	95	7.7	981	12	ADN04717	Adn04717 Antipsori
	9	94.6	7.7	2597	8	AAL51569	Aal51569 Human nuc
	10	92.4	7.5	3639	5	AAS64586	Aas64586 DNA encod
c	11	92	7.4	994	12	ACH87568	Ach87568 Human gen
	12	92	7.4	2873	13	ADR14368	Adr14368 Human NF-
	13	91.8	7.4	1533	12	ADN99064	Adn99064 Novel hum
	14	91.8	7.4	1533	12	ADO00633	Ado00633 Novel hum
	15	91.8	7.4	2110	10	ADB63624	Adb63624 Human cDN
	16	91.8	7.4	2760	10	ADC56695	Adc56695 Human mac
	17	91.8	7.4	3078	10	ADB62468	Adb62468 Human cDN
	18	91	7.4	3755	13	ADR07315	Adr07315 Full leng
	19	91	7.4	4209	13	ADR08223	Adr08223 Full leng
	20	90.4	7.3	2026	8	AAD55863	Aad55863 Human nuc
	21	90.4	7.3	2114	10	ADA53124	Ada53124 Human cod
	22	90.2	7.3	2230	10	ADA52931	Ada52931 Human cod
	23	89.8	7.3	590	12	ACH80103	Ach80103 Human gen
	24	89.2	7.2	2064	10	ADC30762	Adc30762 Human nov
	25	89.2	7.2	2221	11	ADM01344	Adm01344 Human cDN
	26	89.2	7.2	2622	5	AAS68872	Aas68872 DNA encod
	27	89.2	7.2	2729	4	AAH16178	Aah16178 Human cDN
	28	89.2	7.2	3839	6	ABK83826	Abk83826 Human cDN
	29	89.2	7.2	3839	12	ADN04207	Adn04207 Antipsori
	30	89.2	7.2	3839	13	ADR25049	Adr25049 Breast ca
	31	89	7.2	1991	13	ADS09824	Ads09824 Human the
	32	89	7.2	2298	4	AAH16608	Aah16608 Human cDN
	33	89	7.2	2299	13	ADS09825	Ads09825 Human the
	34	89	7.2	2905	5	AAS66143	Aas66143 DNA encod
	35	89	7.2	3003	12	ADN06033	Adn06033 Antipsori
	36	89	7.2	3020	5	AAS92560	Aas92560 DNA encod
	37	89	7.2	3502	4	AAI57845	Aai57845 Human pol
	38	88.6	7.2	1335	12	ADN98906	Adn98906 Novel hum
	39	88.6	7.2	1335	12	ADO00475	Ado00475 Novel hum
	40	88.6	7.2	1466	12	ACH91879	Ach91879 Human gen
	41	88.6	7.2	1875	10	ADI60591	Adi60591 Secreted
	42	88.6	7.2	2597	8	AAD55855	Aad55855 Human nuc
	43	88.6	7.2	4227	9	ACA98938	Aca98938 cDNA enco
c	44	88.4	7.2	1410	12	ACH87337	Ach87337 Human gen
	45	88.4	7.2	1952	10	ADC58104	Adc58104 Zinc fing

ALIGNMENTS

RESULT 1
AAA15398
ID AAA15398 standard; DNA; 2060 BP.
XX
AC AAA15398;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding a transcription factor designated CATFIIIA.
XX
KW Transcription factor; CATFIIIA; DNA-binding protein;
KW ribosomal RNA 5S gene; fungal infection; ss.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT CDS 720. .1958
FT /*tag= a
FT /transl_except= (pos: 1296. .1298, aa: Ser)
FT /transl_except= (pos: 1734. .1736, aa: Ser)
XX
PN WO200028037-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002739.
XX
PR 10-NOV-1998; 98FR-00014147.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Camier S, Sentenac A;
XX
DR WPI; 2000-376549/32.
DR P-PSDB; AAY93316.
XX
PT New nucleic acid encoding Candida albicans transcription factor, useful
PT e.g. in screening for antimycotic agents and for immunization.
XX
PS Claim 4; Page 32-33; 45pp; French.
XX
CC The present sequence encodes a Candida albicans transcription factor,
CC designated CATFIIIA. The polypeptide is a DNA-binding protein, which is

```
CC   involved in initiating transcription of the ribosomal RNA 5S gene. The
CC   polynucleotide is used to screen for its specific inhibitors, potentially
CC   useful as antimycotic agents, to raise an antibody response that is
CC   protective against fungal infection and to raise antibodies. Such
CC   antibodies, as well as the polypeptides and polynucleotides are used in
CC   compositions for diagnosing and treating fungal infections, e.g. by
CC   detecting polymorphisms and mutations
XX
SQ   Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1236;  DB 3;  Length 2060;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-233;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1955


RESULT 2
ABZ32206
ID   ABZ32206 standard; DNA; 1239 BP.
XX
AC   ABZ32206;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential gene SEQ ID NO 6493.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   P-PSDB; ABP73656.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 37; SEQ ID NO 6493; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
```

```
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans gene used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office
XX
SQ   Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

  Query Match             99.7%;  Score 1232.8;  DB 6;  Length 1239;
  Best Local Similarity   99.8%;  Pred. No. 5.5e-233;
  Matches 1234;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236


RESULT 3
ADA53158
ID   ADA53158 standard; cDNA; 2132 BP.
XX
AC   ADA53158;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 726.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54797.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
```

```
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 1; SEQ ID NO 726; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

  Query Match             8.0%;  Score 99.4;  DB 10;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 4.7e-10;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |          || ||  |  |  || |   |      |   | | |
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |     | || || |||||  ||||  ||||| |   |||||  | |||  | |
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||    | ||  || | |    ||   ||| || | ||||       || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |       |||||  | |  |||||||  |||    ||| | | | |
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  ||||||||||   ||||         |||| | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  |  |||||          |    || |||||| |||    |||| |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |      |    | |   ||||| | | | |   | |  ||| |||   | |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||         |||||| ||  |  |||||||  |||
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||     ||    |      ||   |          ||||||| ||  ||
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              |||||  |       ||  ||     |||  | ||    | | | || |
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964


RESULT 4
ACA98970
ID   ACA98970 standard; cDNA; 1705 BP.
XX
AC   ACA98970;
XX
DT   25-JUL-2003  (first entry)
XX
DE   cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX
KW   Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
KW   anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW   antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW   cell proliferative disorder; cancer; atherosclerosis;
KW   neurological disorder; epilepsy; Huntington's disease; stroke;
KW   immune disorder; inflammatory disorder; AIDS; allergy;
KW   developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW   protein-protein interaction; drug-target interaction;
KW   gene expression profile; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003023003-A2.
XX
PD   20-MAR-2003.
XX
PF   05-SEP-2002; 2002WO-US028540.
XX
PR   07-SEP-2001; 2001US-0317792P.
PR   07-SEP-2001; 2001US-0317912P.
PR   14-SEP-2001; 2001US-0322270P.
PR   21-SEP-2001; 2001US-0324040P.
PR   28-SEP-2001; 2001US-0326732P.
PR   19-OCT-2001; 2001US-0346716P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Tang YT,  Jackson JL,  Griffin JA,  Elliott VS,  Forsythe IJ;
PI   Becha SD,  Richardson TW,  Lee EA,  Sprague WW,  Emerling BM;
PI   Thangavelu K,  Warren BA,  Tran UK,  Yue H,  Xu Y,  Yue H,  Li JX;
PI   Hafalia AJA,  Sanjanwala B,  Marquis JP,  Gorvad AE,  Lee SY,  Ison CH;
PI   Baughn MR,  Chawla NK,  Nguyen DB,  Swarnakar A,  Zebarjadian Y,  Shah P;
PI   Thornton M,  Yao MG,  Khan FA,  Gandhi AR,  Yang J,  Kable AE;
PI   Burford N,  Ramkumar J;
XX
DR   WPI; 2003-313243/30.
DR   P-PSDB; ABU96722.
XX
PT   New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT   infections.
XX
PS   Claim 5; Page 340-341; 345pp; English.
XX
CC   The invention describes a novel human isolated nucleic acid-associated
CC   polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC   diagnosing, treating and preventing diseases or conditions associated
CC   with the decreased expression or overexpression of NAAP, such as cell
CC   proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC   epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC   allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC   disorders, or infections. These are also useful in assessing the effects
CC   of exogenous compounds on the expression of nucleic acid and amino acid
CC   sequences of NAAP. The NAAP or its fragments are useful in screening
CC   compounds for effectiveness as agonist or antagonist of the polypeptides,
CC   or in altering the expression of the target polynucleotide and compounds
CC   that specifically bind to or modulate the activity of the polypeptide.
CC   The microarray is useful in monitoring or measuring protein-protein
CC   interactions, drug-target interactions, and gene expression profiles.
CC   This sequence encodes a novel human nucleic acid-associated protein
CC   (NAAP)
XX
SQ   Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

  Query Match             7.8%;  Score 96.8;  DB 9;  Length 1705;
  Best Local Similarity   52.9%;  Pred. No. 1.5e-09;
  Matches  259;  Conservative    0;  Mismatches  222;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||   || ||      |    |   |   |
Db        695 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 754

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | |  | || | | || || |||||  | ||  | ||||| |||  |  | |||    |
Db        755 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 814
```

```
Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  |||||   ||||  || | |   ||   ||| || ||||||| |      || |||
Db        815 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 868

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              | ||||| |||  |   |  |||||  |    | |||||    |   ||| | ||| |
Db        869 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 928

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  ||||| ||||  ||||     |    | | |  | |
Db        929 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 988

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||   |  | ||   |  | || ||| || |||     ||| |||
Db        989 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1048

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
              |  ||     |   | |   | ||||  |||  |  |      || |||   |  |
Db       1049 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAATGCTACAAA 1108

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||||||| |  ||| |||||||  |  || | ||||||  ||
Db       1109 TGTGAAGAATGTGGCAAAGTTTTTAACCGATCTTCAAACCTTACAGACCATAAAAAAATT 1168

Qy        484 CATCATGGTG 493
              |||   || |
Db       1169 CATACAGGAG 1178
```

RESULT 5
ACA56456
ID ACA56456 standard; cDNA; 2320 BP.
XX
AC ACA56456;
XX
DT 06-JUN-2003 (first entry)
XX
DE Human signalling pathway polynucleotide probe SEQ ID NO 1054.
XX
KW Human; probe; ss; array element; Parkinson's disease;
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS Homo sapiens.
XX
PN US6500938-B1.
XX
PD 31-DEC-2002.
XX
PF 30-JAN-1998; 98US-00016434.
XX
PR 30-JAN-1998; 98US-00016434.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Seilhamer JJ;
XX
DR WPI; 2003-352189/33.
XX
PT Combination of polynucleotide probes, useful as array elements in a
PT microarray for monitoring the expression of a number of target
PT polynucleotides.
XX
PS Claim 1; SEQ ID NO 1054; 65pp; English.
XX
CC The invention relates to a combination which, comprises a number of
CC polynucleotide probes comprising a sequence selected from one of the 1490
CC sequences mentioned in the specification. The combination is useful as an
CC array element in a microarray for monitoring the expression of a number
CC of target polynucleotides. The microarray is particularly useful in the
CC diagnosis and treatment of cancer and immunopathology and neuropathology.
CC The microarray is useful in diagnostics and treatment regimens, drug
CC discovery and development, toxicological and carcinogenicity studies,
CC forensics and pharmacogenomics. The microarray is also useful for
CC monitoring progression of diseases and for developing sophisticated
CC profiles for the effects of currently available therapeutic drugs. The
CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC and genomic fragments and in research and diagnostic applications. The
CC array can detect changes in expression in a large number of genes coding
CC for different signaling pathway populations which can be used to diagnose
CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC and Parkinson's disease. The present sequence represents a polynucleotide
CC probe of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=06500938B1
XX
SQ Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.6; DB 10; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.7e-09;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || | |||  |||| |||||| | ||| ||| | |||    |
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                 |||  || ||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||| |||        ||| |  |    |||||||   ||   ||| | | | | ||
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  |  ||||| ||||  ||||          |||| | |  | |||
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              |||  | |  ||||              |  | || |||||| |||  | |||| ||| |
Db       1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy        369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | |       ||   |  | |||| | ||| |   |  |  || ||| | |  |  ||
Db       1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              |||| |||||  ||||| ||| |   |||||   |  | |||||||  ||| ||
Db       1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy        486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
              |  ||| |    |||   |    ||  |          ||||||  | | |  ||
Db       1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy        546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
                | | |    |      |   |    |  | | ||  |   | |||||   |||||
Db       1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy        606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||| |||| || |      |   | |  ||  |||||  |  |||||||   ||
Db       1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551
```

RESULT 6
ADI56252
ID ADI56252 standard; DNA; 2320 BP.
XX
AC ADI56252;
XX
DT 22-APR-2004 (first entry)
XX
DE Human polynucleotide probe #1054.
XX
KW Human; probe; ss; receptor-like polypeptide; transducing polypeptide;

KW effector-like polypeptide; cancer; immunopathology; neuropathology;
KW drug development; toxicology; carcinogenicity;
KW signalling pathway polypeptide; adrenal gland; bladder; bone;
KW bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;
KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;
KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.
XX
OS Homo sapiens.
XX
PN US2004010136-A1.
XX
PD 15-JAN-2004.
XX
PF 26-NOV-2002; 2002US-00305720.
XX
PR 30-JAN-1998; 98US-00016434.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Seilhamer JJ;
XX
DR WPI; 2004-090520/09.
XX
PT New composition comprising polynucleotide probes, useful as array
PT elements in a microarray for monitoring the expression of target
PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic
PT fragments.
XX
PS Claim 6; SEQ ID NO 1054; 73pp; English.
XX
CC The invention relates to a composition of polynucleotide probes
CC comprising first polynucleotide probes comprising at least a portion of a
CC gene encoding a receptor-like polypeptide, second polynucleotide probes
CC comprising at least a portion of a gene encoding a transducing
CC polypeptide and third polynucleotide probes comprising at least a portion
CC of a gene encoding an effector-like polypeptide. The probes of the
CC composition are useful as array elements in a microarray for monitoring
CC the expression of target polynucleotides. The microarray is useful in the
CC diagnosis and treatment of cancer, an immunopathology or a
CC neuropathology. It can also be used for drug discovery and development,
CC toxicological and carcinogenicity studies, forensics or pharmacogenomics.
CC Microarrays can also be used for monitoring the progression of diseases
CC that may be associated with the altered expression of signalling pathway
CC polypeptides. The composition can also be used to purify a subpopulation
CC of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile
CC is also useful for the diagnosis and treatment of cancer, e.g. cancers of
CC the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,
CC an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or
CC ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,
CC epilepsy, Alzheimer's disease or depression. This sequence represents a
CC human polynucleotide probe of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.6; DB 12; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.7e-09;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

Qy 69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db 964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy 129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db 1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy 189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db 1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy 249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db 1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy 309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db 1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy 369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db 1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy 426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db 1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy 486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
Db 1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy 546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
Db 1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy 606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db 1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551

RESULT 7
AAS91317
ID AAS91317 standard; cDNA; 4563 BP.
XX
AC AAS91317;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27121.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27130.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27121; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal

```
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 95.2;  DB 5;  Length 4563;
  Best Local Similarity   52.7%;  Pred. No. 3.5e-09;
  Matches  258;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||         || ||     |  ||     |    ||   | | |
Db       1315 GAATGTGGCAAAGCCTTTAACCAGTCCTCACACCTTACTCAACATAAAACAATTCATACT 1374

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | | | |  | || || |||||  ||||  |||||  | ||| |  |  |||    |
Db       1375 GGAGGGAAAACCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATCTTTCAGCCCTT 1434

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |||||   | ||  || | |   ||   ||| || ||||||      || || |||
Db       1435 ACTAAACATAAGATAATTCATACTGGGGAGAAACCATACAAATGT------GAAGAATGT 1488

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||     |   |||||  |    | |||||  | |   ||| | | | |
Db       1489 GGCAAAGCTTTTAGGCAATCCTCACACCTTACTAGACATAAAGCAATTCATACTGGAGAG 1548

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  ||||| ||||   ||||   |      | || | | ||||
Db       1549 AAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATTTCTCAGACCTTAGAAGA 1608

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||             | | || |||| | |||    |||| |||
Db       1609 CATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGGAAAGCTTTT 1668

Qy        367 TATAAACATCAATCTTTAAGA---CATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                  |  | | |||   |||||  ||| |   |   |   || ||| | |  |
Db       1669 AGCCAGTCCTCAACCCTTAGAAACCATCAGATAATTCATACTGGAGAGAAACCCTACAAA 1728

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||  ||||| ||| |   |  | |||||| |  |   ||||||  |||
Db       1729 TGTGAAGAATGTGGTAAAGCTTTTAAGTGGTCATCAAAACTTACTGTACATAAGGTAATT 1788

Qy        484 CATCATGGTG 493
              |||  ||| |
Db       1789 CATACTGGAG 1798


RESULT 8
ADN04717
ID   ADN04717 standard; cDNA; 981 BP.
XX
AC   ADN04717;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Antipsoriatic cDNA sequence #572.
XX
KW   ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004028479-A2.
XX
PD   08-APR-2004.
XX
PF   25-SEP-2003; 2003WO-US030907.
XX
PR   25-SEP-2002; 2002US-0414006P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Bodary S,  Clark H,  Jackman J,  Schoenfeld J,  Williams PM,  Wood WI;
PI   Wu TD;
XX
DR   WPI; 2004-305105/28.
DR   P-PSDB; ADN04718.
XX
PT   New PRO nucleic acid or polypeptide, useful for preparing a
PT   pharmaceutical composition for diagnosing or treating psoriasis in a
PT   mammal.
XX
PS   Claim 1; SEQ ID NO 1111; 3069pp; English.
XX
CC   The invention relates to novel polynucleotide and polypeptides for
CC   treating psoriasis or a sequence having at least 80% identity to the
CC   above sequences. The nucleic acid is useful for preparing a composition
CC   for diagnosing or treating psoriasis in a mammal. This sequence
CC   corresponds to one of the polynucleotides of the invention.
XX
SQ   Sequence 981 BP; 380 A; 178 C; 170 G; 253 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 95;  DB 12;  Length 981;
  Best Local Similarity   50.4%;  Pred. No. 3.1e-09;
  Matches  289;  Conservative    0;  Mismatches  275;  Indels    9;  Gaps    2;

Qy         89 AAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACA 148
              |||  ||||  |||| || || |  |   |||    |     |||||   | ||  || |
Db          1 AAGAATGTGGCAAAGGCTTTAGTATATTCTCAACCCTTACTAAACATAAGATAATTCATA 60

Qy        149 GTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAAT 208
                |   ||   ||| || ||||| |      | || |||| |||||| |||  | |    |
Db         61 CTGGAGAGAAACCCTACAAATGCA------ATGAATGTGGTAAAGCCTTTAACTGGTCCT 114

Qy        209 CACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTCAGTGT 268
              ||  | |    | |||||     |   ||| | | | | ||||| | | | |||  ||  |
Db        115 CAACTCTTACTAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAAT 174

Qy        269 GTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGT 328
              |||| ||||   ||||       |  | | |  ||||| | ||    ||||
Db        175 GTGGCAAAGCTTTTAACCAGTCCTCAACCCTTACTAGACATAAGATAGTTCATACTGGAG 234

Qy        329 CATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGAC 388
                  |  | || |||| | |||     |||| ||| |          | || ||| ||
Db        235 AGAAACCCTACAAATGTGAAGAATGTGGTAAAGCCTTTAAACGGTCCACAACTCTAAAAC 294

Qy        389 ATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCA 448
              || | | | | | | |  | | || ||| |||   |  ||| || |||||   |||| || |
Db        295 ATAAGAGAATTTATACTAAAGAGAAACCATACAAATGTGAAGAATGTGGAAAAGCCTTTA 354

Qy        449 CTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTTATC 508
              |  |   ||||   |  |  |||||||  |||  |||   ||| | |   |||     |
Db        355 GTGTATTCTCAACCCTTACTAAACATAAGATAATTCATACTGGAGCAAAACCTTACAAAT 414

Qy        509 AATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATA 568
                ||   |    ||  ||||      |||  |||      | | |    |      |
Db        415 GTGACGAATGTGGCAGTGCCTTTAGGGCATTCTCAACCCTTACTGAACATAAGAGAGTTC 474

Qy        569 TAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAA 628
                |    | | | ||  |    |||||   | ||||||||||||  | |    ||
Db        475 ATACTGGAGAGAAACCTTAC---AAATGCAATGAATGTGGTAAAGCCTTTAACTGGTCCT 531

Qy        629 AAGGTTTATCTTCACATATGTTAAGTCATGATG 661
                | | |  ||  ||||| |  || ||||  ||
Db        532 CAACTCTTACTAAACATAAGAGAATTCATACTG 564
```

```
RESULT 9
AAL51569
ID   AAL51569 standard; DNA; 2597 BP.
XX
AC   AAL51569;
XX
DT   10-APR-2003  (first entry)
XX
DE   Human nucleic acid-associated protein coding sequence - SEQ ID No 52.
XX
KW   Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW   cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW   cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW   mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW   Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW   Crohn's disease; transgenic animal; animal model.
XX
OS   Homo sapiens.
XX
PN   WO2003000864-A2.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-US021179.
XX
PR   22-JUN-2001; 2001US-0300518P.
PR   29-JUN-2001; 2001US-0301787P.
PR   29-JUN-2001; 2001US-0301792P.
PR   29-JUN-2001; 2001US-0301892P.
PR   29-JUN-2001; 2001US-0301893P.
PR   06-JUL-2001; 2001US-0303405P.
PR   06-JUL-2001; 2001US-0303442P.
PR   15-MAR-2002; 2002US-0364438P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Gandhi AR,  Swarnakar A,  Hafalia AJA,  Warren BA,  Emerling BM;
PI   Arvizu CS,  Ison CH,  Honchell CD,  Lee EA,  Yue H,  Forsythe IJ;
PI   Ramkumar J,  Griffin JA,  Yang J,  Sanjanwala MM,  Baughn MR;
PI   Borowsky ML,  Yao MG,  Walia NK,  Bandman O,  Lal PG,  Becha SD,  Lee SY;
PI   Richardson TW,  Elliott VS,  Luo W,  Tang YT,  Zebarjadian Y,  Lu Y;
XX
DR   WPI; 2003-201420/19.
DR   P-PSDB; AAO16419.
XX
PT   New nucleic acid-associated proteins and polynucleotides, useful for
PT   diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT   neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT   disorders (e.g. AIDS).
XX
PS   Claim 12; Page 292-293; 312pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of human
CC   nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC   the invention are useful for diagnosing, treating or preventing disorders
CC   associated with aberrant expression of NAAP, such as: cell proliferative
CC   disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC   or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC   or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC   Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC   (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC   the invention are useful for creating transgenic animals to model human
CC   disease. The present DNA sequence encodes a human nucleic acid-associated
CC   protein of the invention
XX
SQ   Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 94.6;  DB 8;  Length 2597;
  Best Local Similarity   55.1%;  Pred. No. 4.2e-09;
  Matches  231;  Conservative    0;  Mismatches  179;  Indels    9;  Gaps    2;

Qy         78 TTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTT 137
              ||  | || |||||  ||||  ||||||| ||||    | |||    |    ||||||
Db       1070 TTATAAATGTGAAGAATGTGGCAAAGCCTTTAATGTGTCCTCAACCCTTACTCAACATAA 1129

Qy        138 AAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATT 197
               ||||  || | |   ||   ||| || ||||||      || || ||||  ||||| ||
Db       1130 GAGAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAGTGTGGCAAAGCCTT 1183

Qy        198 TTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCA 257
              |  |      |||  | |       |||||     |   ||| | | | |||||||||  | |
Db       1184 TAACGTGTCCTCAACTCTTACTCAACATAAGAGAATTCATACTGGAGAAAAACCATACAA 1243

Qy        258 TTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCAC 317
              |||  ||  ||||| ||||   ||||  |      |||| | |  | |||| |||  |
Db       1244 ATGTGAAGAATGTGGCAAAGCCTTTAACACATCCTCACACCTCACCACACATAAAAGAAT 1303

Qy        318 CCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACA 374
              |||||              | | || |||| | |||    |||| ||| | |
Db       1304 TCATACCGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTTCTC 1363

Qy        375 TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATG 434
              |||  |   | |||| | ||| |   |  |   || ||| | |  |  ||||||  ||||
Db       1364 ACAACTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAGAATG 1423

Qy        435 TAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTG 493
              |  |||| ||| |  ||  | |||||  |  |  ||||||   |||  |||  ||| |
Db       1424 TGGCAAAGCTTTTAAGCGGTCCTCAAACCTTACTGAACATAGGATAATTCATACTGGAG 1482


RESULT 10
AAS64586
ID   AAS64586 standard; cDNA; 3639 BP.
XX
AC   AAS64586;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #390.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG00399.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 390; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
```

```
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 92.4;  DB 5;  Length 3639;
  Best Local Similarity   54.1%;  Pred. No. 1.2e-08;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||  |    | |||    | |
Db        738 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 797

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db        798 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 851

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               ||||| |||  | |    |||    | |  | |||||    |   ||| | | | ||
Db        852 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 911

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  ||||| ||||   |||    |     | | |   | ||
Db        912 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 971

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |          | | || | || | ||| |  |||| ||| |
Db        972 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAGCTTTTAA 1031

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                 ||    | |   |   | |||| | ||| ||  || |   || |||   |  |  ||
Db       1032 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1091

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |  |   ||||| || |  ||
Db       1092 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1151

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db       1152 TACTGGAGAGAAACCT 1167


RESULT 11
ACH87568/c
ID   ACH87568 standard; DNA; 994 BP.
XX
AC   ACH87568;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon probe #20763.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 1; SEQ ID NO 20763; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 994 BP; 252 A; 159 C; 203 G; 380 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 92;  DB 12;  Length 994;
  Best Local Similarity   52.0%;  Pred. No. 1.2e-08;
  Matches  259;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||    |      || ||      |  || |   |      |   | | |
Db        763 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 704

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | ||| |    | || || |||||  |||| |||||| | ||| |  || |||    |
Db        703 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 644

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                  ||||      |  ||   ||| ||  ||| || ||||||      || || |||
Db        643 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 590
```

```
Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |    | |||||   ||   ||||| | | |
Db        589 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 530

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | ||| |   ||||| |||||  |||| |        | | | | || |
Db        529 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 470

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| |||  |  |||||          |  | || |||| | |||    |||| |||
Db        469 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 410

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               ||     |   ||   |  | |||| | ||  |   | |   || ||| | | |
Db        409 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 350

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |  |   | ||||| |  |   ||||| |||
Db        349 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 290

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |     |||
Db        289 CATACTGGAGAGAAACCT 272
```

RESULT 12
ADR14368
ID ADR14368 standard; DNA; 2873 BP.
XX
AC ADR14368;
XX
DT 21-OCT-2004 (first entry)
XX
DE Human NF-kappaB pathway-associated gene SeqID369.
XX
KW NF-kappaB pathway; antiinflammatory; cytostatic; hepatotropic; virucide;
KW antiarthritic; antirheumatic; gastrointestinal-Gen; antiasthmatic;
KW antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;
KW immunosuppressive; vulnerary; gene therapy; immune disorder;
KW inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;
KW hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;
KW hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;
KW X-linked anhidrotic ectodermal dysplasia; immunodeficiency;
KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;
KW viral replication; host cell survival; evasion of immune response;
KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;
KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;
KW autoimmune disorder; hyper immune activity;
KW aberrant acute phase response; hypercongenital condition; birth defect;
KW necrotic lesion; wound; organ transplant rejection;
KW aberrant signal transduction; proliferating disorder; cancer;
KW HIV propagation; gene; ds; human.
XX
OS Homo sapiens.
XX
PN WO2004065577-A2.
XX
PD 05-AUG-2004.
XX
PF 13-JAN-2004; 2004WO-US000798.
XX
PR 14-JAN-2003; 2003US-0440068P.
PR 12-MAY-2003; 2003US-0469757P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Nadler SG, Neubauer MG, Feder JN, Carman J;
XX
DR WPI; 2004-562168/54.
DR P-PSDB; ADR14369.
XX
PT New isolated polynucleotides and polypeptides associated with NF-kappaB
PT pathway, useful for diagnosing, treating, or preventing disorders or
PT diseases associated with NF-kappaB pathway.
XX
PS Claim 1; SEQ ID NO 369; 237pp; English.
XX
CC This invention relates to the novel association of protein sequences (and
CC the genes which encode them) to the NF-kappaB pathway. The invention may
CC be useful for the production of compounds with an antiinflammatory,
CC cytostatic, hepatotropic, virucide, antiarthritic, antirheumatic,
CC gastrointestinal-Gen, antiasthmatic, antiarteriosclerotic,
CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or
CC vulnerary activity or for gene therapy. The proteins and nucleotides are
CC useful for diagnosing, preventing, treating, or ameliorating conditions
CC or diseases associated with the NF-kappaB pathway. The condition is an
CC immune disorder, an inflammatory disorder, an inflammatory disorder
CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,
CC hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM
CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic
CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,
CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell
CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory
CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick
CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper
CC immune activity, disorders related to aberrant acute phase responses,
CC hypercongenital conditions, birth defects, necrotic lesions, wounds,
CC organ transplant rejection, conditions related to organ transplant
CC rejection, disorders related to aberrant signal transduction,
CC proliferating disorders, cancers and HIV propagation in cells infected
CC with other viruses. The present sequence is that of a human gene which is
CC subject to the novel association with the NF-kappaB pathway of the
CC invention. Note: This sequence does not appear in the specification but
CC was obtained by the indexer from Genbank.
XX
SQ Sequence 2873 BP; 948 A; 564 C; 545 G; 816 T; 0 U; 0 Other;

Query Match 7.4%; Score 92; DB 13; Length 2873;
Best Local Similarity 52.2%; Pred. No. 1.4e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||     || ||     |      |   |   |
Db       1621 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 1680

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              |  |  | || | |  || || |||||  | ||  | |||||  |||   |  | |||    |
Db       1681 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 1740

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |  |||||   ||||  || | |   ||   ||| || |||||| |       || |||
Db       1741 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 1794

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |   |  |||||  |    | |||||    |   ||| | ||| |
Db       1795 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 1854

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  |||| ||||   |||       |     | | |  | |
Db       1855 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 1914

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||   |  | ||   |  | || ||| || |||     ||| |||
Db       1915 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1974

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
              |  ||     |    | |   | ||||  |||  |  |      || |||   |
Db       1975 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 2034

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||    |||| ||| |  ||| | |||||| |  |  |||||| | |
Db       2035 TGTGAAGAATGCGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGTACATT 2094

Qy        484 CATCATGGTG 493
              |||  ||| |
```

```
Db       2095 CATACTGGAG 2104


RESULT 13
ADN99064
ID   ADN99064 standard; cDNA; 1533 BP.
XX
AC   ADN99064;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Novel human cDNA sequence #664.
XX
KW   ds; gene; anti-inflammatory; dermatological; neuroprotective;
KW   immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;
KW   gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW   psoriasis; diabetes; early aging; hormonal imbalance;
KW   ischemic heart disease; ulcerative colitis.
XX
OS   Homo sapiens.
XX
PN   WO2004038003-A2.
XX
PD   06-MAY-2004.
XX
PF   24-OCT-2003; 2003WO-US033947.
XX
PR   25-OCT-2002; 2002US-0421061P.
PR   25-OCT-2002; 2002US-0421080P.
PR   25-OCT-2002; 2002US-0421552P.
PR   25-OCT-2002; 2002US-0421614P.
PR   30-OCT-2002; 2002US-0422177P.
PR   30-OCT-2002; 2002US-0422178P.
PR   15-NOV-2002; 2002US-0426355P.
PR   15-NOV-2002; 2002US-0426384P.
PR   15-NOV-2002; 2002US-0426394P.
PR   15-NOV-2002; 2002US-0426430P.
PR   15-NOV-2002; 2002US-0426916P.
PR   27-NOV-2002; 2002US-0429224P.
PR   27-NOV-2002; 2002US-0429275P.
PR   27-NOV-2002; 2002US-0429302P.
PR   27-NOV-2002; 2002US-0429326P.
PR   27-NOV-2002; 2002US-0429651P.
PR   04-DEC-2002; 2002US-0430645P.
PR   04-DEC-2002; 2002US-0430651P.
PR   04-DEC-2002; 2002US-0430657P.
PR   04-DEC-2002; 2002US-0430663P.
PR   04-DEC-2002; 2002US-0430668P.
PR   04-DEC-2002; 2002US-0430684P.
PR   05-DEC-2002; 2002US-0430937P.
PR   05-DEC-2002; 2002US-0430965P.
PR   05-DEC-2002; 2002US-0431458P.
PR   12-DEC-2002; 2002US-0433251P.
PR   12-DEC-2002; 2002US-0433500P.
PR   13-DEC-2002; 2002US-0433316P.
PR   13-DEC-2002; 2002US-0433318P.
PR   23-DEC-2002; 2002US-0436236P.
PR   03-JAN-2003; 2003US-0437914P.
PR   17-JAN-2003; 2003US-0440820P.
PR   17-JAN-2003; 2003US-0440821P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476621P.
PR   09-JUN-2003; 2003US-0476632P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485217P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   08-JUL-2003; 2003US-0485359P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Kothakota S,  Lin H,  Linnemann T,  Pierce K,  Wang Y;
PI   Wong JGP,  Wu G,  Zhang H,  Zeng C;
XX
DR   WPI; 2004-365511/34.
DR   P-PSDB; ADN99848.
XX
PT   New nucleic acid molecules, useful in preparing a composition for
PT   treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT   disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT   ulcerative colitis.
XX
PS   Claim 1; SEQ ID NO 664; 532pp; English.
XX
CC   The invention relates to a nucleic acid molecule comprising a
CC   polynucleotide sequence or its complement that encodes a polypeptide. The
CC   nucleic acid is useful in preparing a composition for treating or
CC   preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC   cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC   heart disease or ulcerative colitis. This sequence corresponds to a
CC   nucleic acid of the invention.
XX
SQ   Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 91.8;  DB 12;  Length 1533;
  Best Local Similarity   52.1%;  Pred. No. 1.4e-08;
  Matches  257;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy         11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
              ||| | || |     |  ||    ||   | || ||      |      |   | | | | |
Db        305 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 364

Qy         71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
              |||||  || || || |||||  |||| |||||||| |||  |  | |||     |
Db        365 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 424

Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
              |||||   | ||  || | |   ||   ||| || ||||||      || || ||||  |
Db        425 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 478

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
              |||| |||  | |    ||| || |    | |||||   ||   ||| | | | | ||||
Db        479 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 538

Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | | |||  ||  ||||| ||||   ||||       |  | | |  | ||||
Db        539 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 598

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  |||||| |          |  | || |||| | |||    | || ||| |
Db        599 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 658

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
```

```
Db        659 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 718

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487

Db        719 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 778

Qy        488 ATGGTGGATCTCC 500

Db        779 CTGGGGAAAAACC 791
```

RESULT 14
ADO00633
ID ADO00633 standard; cDNA; 1533 BP.
XX
AC ADO00633;
XX
DT 29-JUL-2004 (first entry)
XX
DE Novel human cDNA sequence #1448.
XX
KW ds; gene; anti-inflammatory; dermatological; neuroprotective;
KW immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;
KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW psoriasis; diabetes; early aging; hormonal imbalance;
KW ischemic heart disease; ulcerative colitis.
XX
OS Homo sapiens.
XX
PN WO2004038003-A2.
XX
PD 06-MAY-2004.
XX
PF 24-OCT-2003; 2003WO-US033947.
XX
PR 25-OCT-2002; 2002US-0421061P.
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421552P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422177P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 15-NOV-2002; 2002US-0426394P.
PR 15-NOV-2002; 2002US-0426430P.
PR 15-NOV-2002; 2002US-0426916P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 27-NOV-2002; 2002US-0429302P.
PR 27-NOV-2002; 2002US-0429326P.
PR 27-NOV-2002; 2002US-0429651P.
PR 04-DEC-2002; 2002US-0430645P.
PR 04-DEC-2002; 2002US-0430651P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430668P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430937P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433251P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433316P.
PR 13-DEC-2002; 2002US-0433318P.
PR 23-DEC-2002; 2002US-0436236P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 17-JAN-2003; 2003US-0440821P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463708P.
PR 18-APR-2003; 2003US-0463716P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 19-MAY-2003; 2003US-0471336P.
PR 22-MAY-2003; 2003US-0472420P.
PR 22-MAY-2003; 2003US-0472430P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476621P.
PR 09-JUN-2003; 2003US-0476632P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485217P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;
PI Wong JGP, Wu G, Zhang H, Zeng C;
XX
DR WPI; 2004-365511/34.
DR P-PSDB; ADN99848.
XX
PT New nucleic acid molecules, useful in preparing a composition for
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT ulcerative colitis.
XX
PS Claim 1; SEQ ID NO 2232; 532pp; English.
XX
CC The invention relates to a nucleic acid molecule comprising a
CC polynucleotide sequence or its complement that encodes a polypeptide. The
CC nucleic acid is useful in preparing a composition for treating or
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
CC nucleic acid of the invention.
XX
SQ Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

  Query Match   7.4%; Score 91.8; DB 12; Length 1533;
  Best Local Similarity   52.1%; Pred. No. 1.4e-08;
  Matches 257; Conservative   0; Mismatches 227; Indels   9; Gaps   2;

```
Qy         11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70

Db        305 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 364

Qy         71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130

Db        365 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 424

Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190

Db        425 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 478

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250

Db        479 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 538
```

```
Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | | |||  ||  ||||| ||||   ||||         |  | | |  | ||||
Db        539 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 598

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  ||||||  |           |  | || |||| | |||    | || ||| |
Db        599 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 658

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
              |       | |    | |||| | | | |   | |  | || ||| | |  |  |||
Db        659 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 718

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
              || |||||   ||||  || |    |   ||||   |   | |||||||  |||  |||
Db        719 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 778

Qy        488 ATGGTGGATCTCC 500
               ||| | |   ||
Db        779 CTGGGGAAAAACC 791


RESULT 15
ADB63624
ID   ADB63624 standard; cDNA; 2110 BP.
XX
AC   ADB63624;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding clone THYMU20071120.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             170. .1792
FT                   /*tag=  a
FT                   /product= "Clone THYMU20071120 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   P-PSDB; ADB65594.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a cDNA of the invention. Note: Some of the sequence
CC   data for this patent is not represented in the printed specification, but
CC   is based on sequence information supplied by the European Patent Office.
XX
SQ   Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 91.8;  DB 10;  Length 2110;
  Best Local Similarity   51.7%;  Pred. No. 1.5e-08;
  Matches  314;  Conservative    0;  Mismatches  277;  Indels   16;  Gaps    4;

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              | | || || || |    | || || | |||  ||||  ||||| | ||| |   | |||
Db       1253 CATACCGAAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAAACACTCCTCA 1312

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                 |    ||||  ||||  |||| |   ||   ||| || ||||||      ||
Db       1313 GCCCTTACTACACATAAGAGAATTCACACTGGAGAGAAACCCTACAAATGT------GAA 1366

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              || |||||  ||||| |||  | ||   ||| | |      |||||    |   ||| |
Db       1367 GAATGTGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGAAACTTCATACT 1426

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              | ||| ||||| | | | |||  ||  ||||| ||||  |||          | | |
Db       1427 GGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTATCCAATCCTCAAAACTT 1486

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |    |||| |||  |  ||| |         |    |  |  | |||| | |||    ||
Db       1487 ACTGAACATAAAAAAATTCATTCTGGAGAGATACCCTACAAGTGTGAAGAATGTGGCAAA 1546

Qy        361 GCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAAAACA 417
              || ||| |  |       |||  |    | |||| | | | |   |  |   || ||| |
Db       1547 GCTTTTAAACACTCCTCATCCCTTACTACACATAAAAGAATTCATACTGGGGAGAAACCC 1606

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              |  |  ||| || |||||   |||| ||| |  ||| | |||||| |  |  |||||||
Db       1607 TACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGATCCTCAAAACTTACTGAACATAAG 1666

Qy        478 TTAAAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
              |||  |||  ||| | ||   |||    ||     | |||    |       ||| |  ||
Db       1667 ATAATTCATACTGGAGAGAAACCCTATAAATGTGAGAGATGTGACAAAGCTTTTAACCAA 1726

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
              | | ||||| |       || || |  |  | | | |   |||| |   |||    | |
Db       1727 TCTGCAAACCT------TACTAAACATAAAAAAATACATACTGGAGAGAAACTACAGAAC 1780

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
              |  |  |||| ||| | || ||     |   |  ||| |||  | |||    || |||
Db       1781 TGGAATGTGTGATAATGATTTTGACAACACCTCAAATTTTTCTAAATATAAAGGAAATCA 1840

Qy        657 TGATGAT 663
              |  || |
Db       1841 TACTGGT 1847


Search completed: February  9, 2005, 02:39:49
Job time : 647.722 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:51:00 ; Search time 223.728 Seconds
(without alignments)
9039.698 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	581.8	47.1	627	4	US-09-248-796A-5184	Sequence 5184, Ap
	2	391	31.6	462	4	US-09-248-796A-5186	Sequence 5186, Ap
	3	96.6	7.8	2320	4	US-09-016-434-1054	Sequence 1054, Ap
	4	90.8	7.3	3798	4	US-09-949-016-4204	Sequence 4204, Ap
	5	90.8	7.3	156942	4	US-09-949-016-12227	Sequence 12227, A
	6	90.8	7.3	156950	4	US-09-949-016-15946	Sequence 15946, A
	7	89.2	7.2	3839	4	US-09-949-016-485	Sequence 485, App
c	8	86.8	7.0	601	4	US-09-949-016-27359	Sequence 27359, A
c	9	86.8	7.0	601	4	US-09-949-016-149638	Sequence 149638,
	10	86.2	7.0	601	4	US-09-949-016-44799	Sequence 44799, A
	11	86.2	7.0	45138	4	US-09-949-016-13027	Sequence 13027, A
c	12	85.6	6.9	601	4	US-09-949-016-27360	Sequence 27360, A
c	13	85.6	6.9	601	4	US-09-949-016-149637	Sequence 149637,
c	14	83.2	6.7	601	4	US-09-949-016-27361	Sequence 27361, A
c	15	83.2	6.7	601	4	US-09-949-016-149636	Sequence 149636,
	16	81.6	6.6	472	4	US-09-513-999C-12522	Sequence 12522, A
	17	80.2	6.5	2804	4	US-09-949-016-2278	Sequence 2278, Ap
	18	80.2	6.5	22294	4	US-09-949-016-14020	Sequence 14020, A
	19	79.4	6.4	1030	4	US-09-976-594-370	Sequence 370, App
	20	79.2	6.4	2982	4	US-09-949-016-4184	Sequence 4184, Ap
	21	79.2	6.4	41435	4	US-09-949-016-15926	Sequence 15926, A
	22	78.2	6.3	3400	4	US-09-799-451-136	Sequence 136, App
c	23	77.2	6.2	601	4	US-09-949-016-149473	Sequence 149473,
c	24	76	6.1	601	4	US-09-949-016-149474	Sequence 149474,
	25	76	6.1	2501	4	US-09-949-016-5363	Sequence 5363, Ap
	26	76	6.1	2504	4	US-09-949-016-487	Sequence 487, App
	27	76	6.1	29671	4	US-09-949-016-12229	Sequence 12229, A
	28	76	6.1	29671	4	US-09-949-016-17105	Sequence 17105, A
	29	74.4	6.0	2143	4	US-09-881-578A-3	Sequence 3, Appli
	30	73	5.9	2925	4	US-09-620-312D-163	Sequence 163, App
	31	71.2	5.8	979	4	US-09-976-594-939	Sequence 939, App
	32	69.6	5.6	2402	4	US-09-949-016-1710	Sequence 1710, Ap
	33	69.6	5.6	2407	4	US-09-949-016-489	Sequence 489, App
	34	69.6	5.6	30867	4	US-09-949-016-12231	Sequence 12231, A
	35	69.6	5.6	30868	4	US-09-949-016-13452	Sequence 13452, A
	36	68	5.5	2891	4	US-09-949-016-1115	Sequence 1115, Ap
c	37	67.6	5.5	601	4	US-09-949-016-39838	Sequence 39838, A
c	38	67.6	5.5	601	4	US-09-949-016-160861	Sequence 160861,
c	39	66.8	5.4	18773	4	US-09-949-016-14164	Sequence 14164, A
	40	66.4	5.4	2891	4	US-09-949-016-4503	Sequence 4503, Ap
	41	66.4	5.4	21360	4	US-09-949-016-12857	Sequence 12857, A
	42	66.4	5.4	21361	4	US-09-949-016-16245	Sequence 16245, A
	43	66.2	5.4	2752	4	US-09-949-016-3563	Sequence 3563, Ap
	44	66.2	5.4	18824	4	US-09-949-016-15305	Sequence 15305, A
	45	65.4	5.3	2274	4	US-09-620-312D-154	Sequence 154, App

ALIGNMENTS

RESULT 1
US-09-248-796A-5184
; Sequence 5184, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5184
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5184

Query Match 47.1%; Score 581.8; DB 4; Length 627;
Best Local Similarity 97.4%; Pred. No. 1.1e-123;
Matches 605; Conservative 0; Mismatches 7; Indels 9; Gaps 1;

```
Qy        616 TGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATC 675
              ||| | || ||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         13 TGTTTGGGAAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTTACCATGATC 72

Qy        676 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 735
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 132

Qy        736 GAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAA 795
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        133 GAACATTATAATATCTTCCATGATGGTAATATCCCCGATGATTTATTAAAGGAAACTGAA 192

Qy        796 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 252

Qy        856 GAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAAGATAGTCTAGATGAA 915
              |||||||||||||||||||          ||||||||||||||||||||||||||||||
Db        253 GAAACAGAGAAATTAAAAGTG---------GAAGAAGATGAAGAAGATAGTCTAGATGAA 303

Qy        916 AAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 975
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        304 AGAAGAAGTGATGTTAGATCAGATTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 363
```

```
Qy        976 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 423

Qy       1036 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 483

Qy       1096 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 543

Qy       1156 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 603

Qy       1216 GAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||
Db        604 GAAACATCAGTGATTTCTCGA 624
```

RESULT 2
US-09-248-796A-5186
; Sequence 5186, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5186
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5186

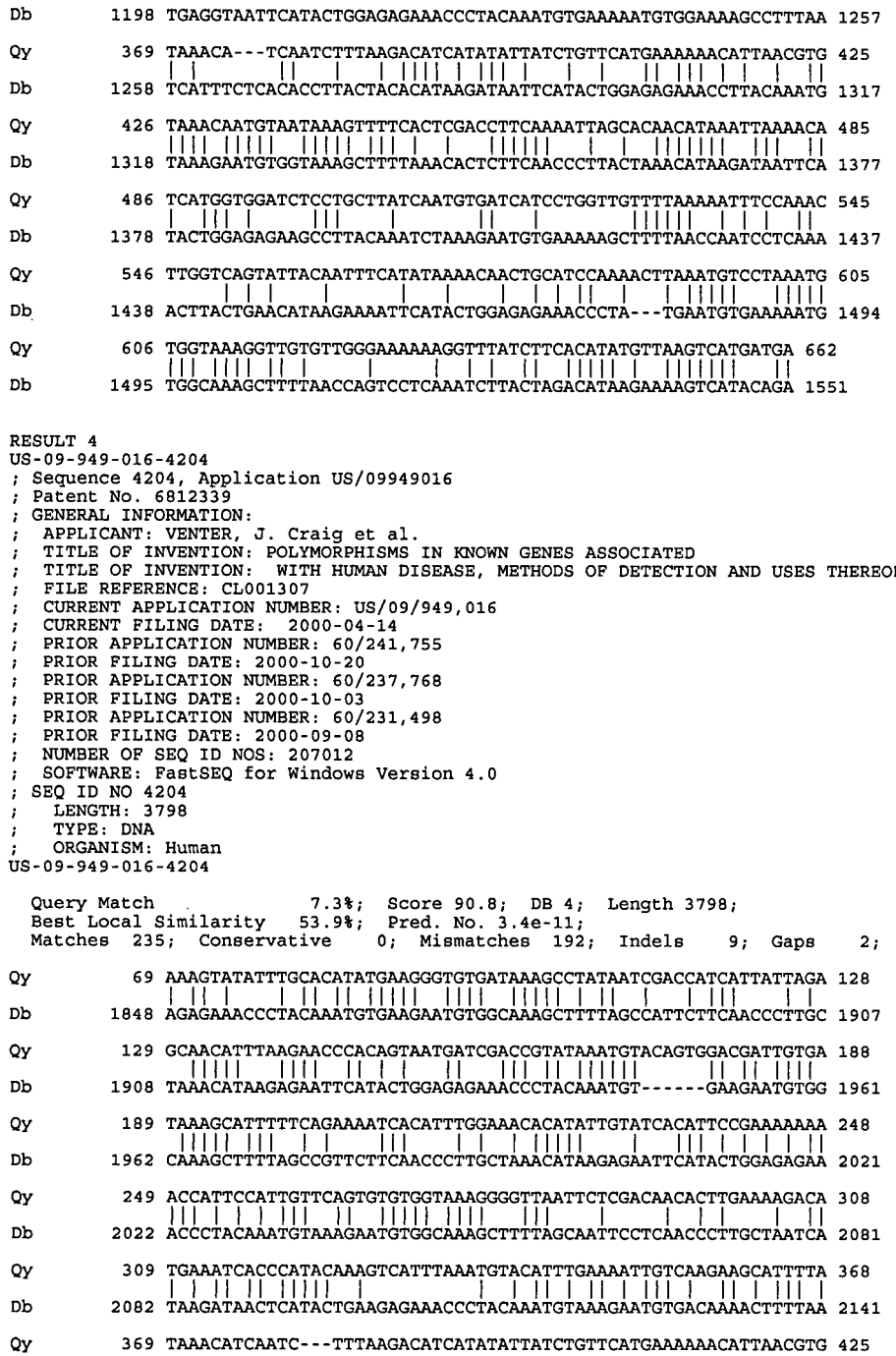
Query Match 31.6%; Score 391; DB 4; Length 462;
Best Local Similarity 98.5%; Pred. No. 4.1e-80;
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         52 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 111

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        112 CGTCCCAAAAAGTATATTTGTACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 171

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 231

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 291

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 351

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        352 AAAAGACATGAAATCACCCATACCAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 411

Qy        361 GCATTTTAT-AAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410
              |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        412 ACATTTTATAAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 462
```

RESULT 3
US-09-016-434-1054
; Sequence 1054, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1054:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1017721
US-09-016-434-1054

Query Match 7.8%; Score 96.6; DB 4; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.4e-12;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   | || || | |||  |||| |||||| | ||| ||| | |||    |
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                  |||  || ||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||| |||        ||| |  |   |||||||   ||   ||| | | | ||
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  |   |||| ||||   ||||          |||| | |  | |||
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              |||  | |  |||||             |  | || |||||| |||  | |||| ||| |
```



RESULT 4
US-09-949-016-4204
; Sequence 4204, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4204
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4204

Query Match 7.3%; Score 90.8; DB 4; Length 3798;
Best Local Similarity 53.9%; Pred. No. 3.4e-11;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

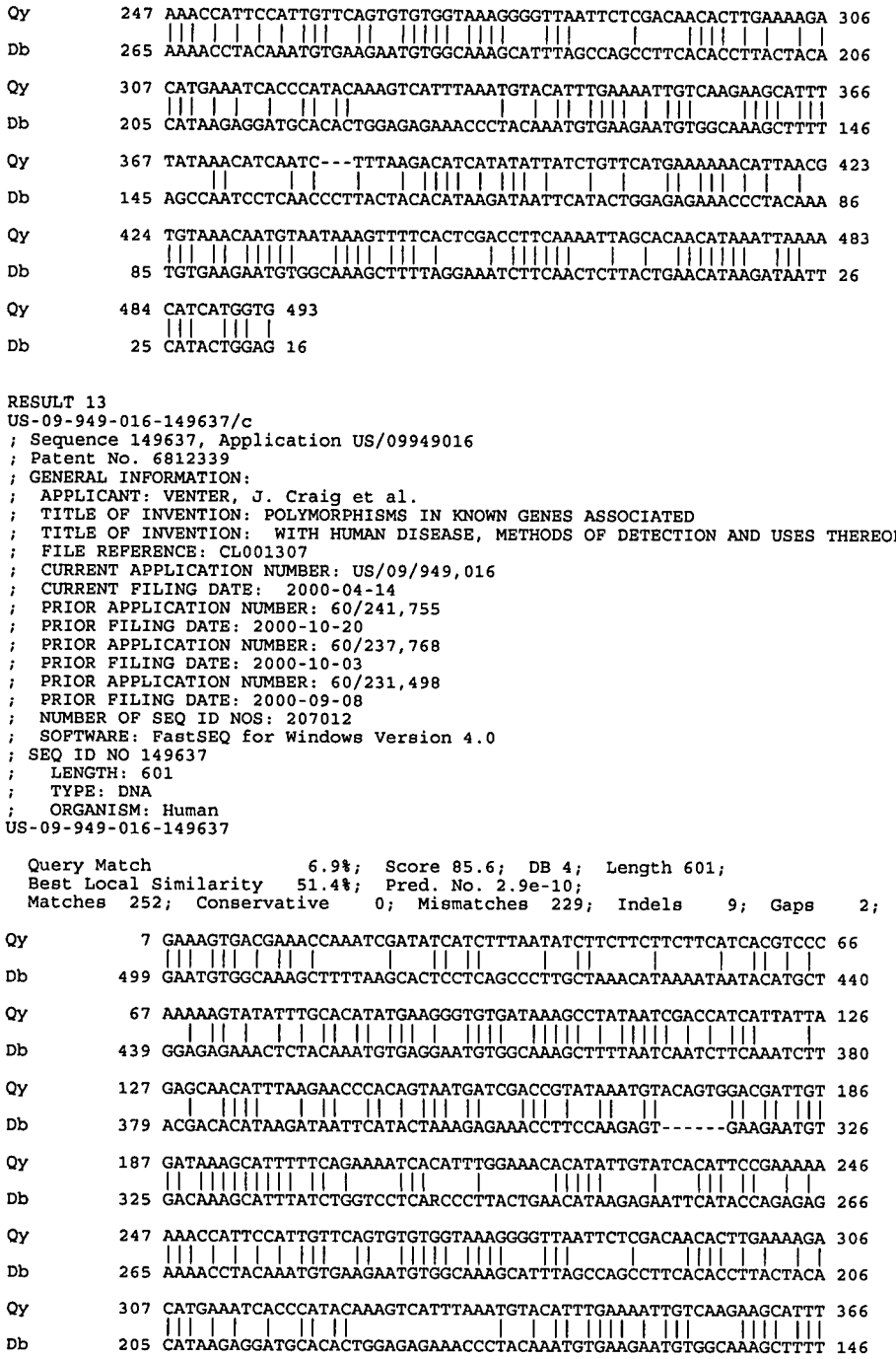
RESULT 5
US-09-949-016-12227
; Sequence 12227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12227
; LENGTH: 156942
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(156942)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12227

Query Match 7.3%; Score 90.8; DB 4; Length 156942;
Best Local Similarity 53.9%; Pred. No. 1.1e-10;
Matches 235; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

```
Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db     153413 TACTGGAGAGAAACCT 153428


RESULT 6
US-09-949-016-15946
; Sequence 15946, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15946
;   LENGTH: 156950
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(156950)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15946

  Query Match             7.3%;  Score 90.8;  DB 4;  Length 156950;
  Best Local Similarity   53.9%;  Pred. No. 1.1e-10;
  Matches  235;  Conservative    0;  Mismatches  192;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  |||||  | ||  |    | |||    | |
Db     152999 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 153058

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db     153059 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 153112

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  | |    |||    | |  | |||||    |   ||| | | | | ||
Db     153113 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 153172

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | | | |||  ||  ||||| ||||   |||    |    | | |    | ||
Db     153173 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 153232

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |          | | || | || | ||| |  || | ||| |
Db     153233 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 153292

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                ||    | |   |   | |||| | ||| ||  || |   || |||   |  |  ||
Db     153293 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 153352

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |  |   |||||| || |  ||
Db     153353 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 153412

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db     153413 TACTGGAGAGAAACCT 153428


RESULT 7
US-09-949-016-485
; Sequence 485, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 485
;   LENGTH: 3839
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-485

  Query Match             7.2%;  Score 89.2;  DB 4;  Length 3839;
  Best Local Similarity   53.7%;  Pred. No. 7.9e-11;
  Matches  234;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  |||||  | ||  |    | |||    | |
Db        964 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  | |    |||    | |  | |||||    |   ||| | | | | ||
Db       1078 CAAAGCTTTTAGCCATTCTTCAGCCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | | | |||  ||  ||||| ||||   |||    |    | | |    | ||
Db       1138 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |          | | || | || | ||| |  || | ||| |
Db       1198 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 1257

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                ||    | |   |   | |||| | ||| ||  || |   || |||   |  |  ||
Db       1258 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |  |   |||||| || |  ||
Db       1318 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1377

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db       1378 TACTGGAGAGAAACCT 1393


RESULT 8
US-09-949-016-27359/c
; Sequence 27359, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
```

```
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27359
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-27359

  Query Match             7.0%;  Score 86.8;  DB 4;  Length 601;
  Best Local Similarity   51.4%;  Pred. No. 1.6e-10;
  Matches  252;  Conservative    1;  Mismatches  228;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |       |   || ||      |  ||      |     | || | |
Db        538 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 479

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||    |
Db        478 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 419

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |  ||||   | ||  || | ||| ||   ||| |  ||  ||      || || |||
Db        418 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 365

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||||||| || |   |||     |       |||||    |   ||| ||  | |
Db        364 GACAAAGCATTTATCTGGTCCTCAACCCTTACTGAACATAAGAGAATTCATACCAGAGAG 305

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||:| | | | |||  ||  ||||| ||||   |||      |    |||| | |  | |
Db        304 AAAMCCTACAAATGTGAAGAATGTGGCAAAGCATTTAGCCAGCCTTCACACCTTACTACA 245

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |   |  || ||                | | || |||| | |||    |||| |||
Db        244 CATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTT 185

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                ||      | |   |   | |||| | ||| |   |  |   || ||| | |  |
Db        184 AGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAA 125

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| ||| |    | ||||||   |  |  ||||||| |||
Db        124 TGTGAAGAATGTGGCAAAGCTTTTAGGAAATCTTCAACTCTTACTGAACATAAGATAATT 65

Qy        484 CATCATGGTG 493
              |||  ||| |
Db         64 CATACTGGAG 55


RESULT 9
US-09-949-016-149638/c
; Sequence 149638, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149638
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-149638

  Query Match             7.0%;  Score 86.8;  DB 4;  Length 601;
  Best Local Similarity   51.4%;  Pred. No. 1.6e-10;
  Matches  252;  Conservative    1;  Mismatches  228;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |       |   || ||      |  ||      |     | || | |
Db        538 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 479

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||    |
Db        478 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 419

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |  ||||   | ||  || | ||| ||   ||| |  ||  ||      || || |||
Db        418 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 365

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||||||| || |   |||     |       |||||    |   ||| ||  | |
Db        364 GACAAAGCATTTATCTGGTCCTCAACCCTTACTGAACATAAGAGAATTCATACCAGAGAG 305

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||:| | | | |||  ||  ||||| ||||   |||      |    |||| | |  | |
Db        304 AAAMCCTACAAATGTGAAGAATGTGGCAAAGCATTTAGCCAGCCTTCACACCTTACTACA 245

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |   |  || ||                | | || |||| | |||    |||| |||
Db        244 CATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTT 185

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                ||      | |   |   | |||| | ||| |   |  |   || ||| | |  |
Db        184 AGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAA 125

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| ||| |    | ||||||   |  |  ||||||| |||
Db        124 TGTGAAGAATGTGGCAAAGCTTTTAGGAAATCTTCAACTCTTACTGAACATAAGATAATT 65

Qy        484 CATCATGGTG 493
              |||  ||| |
Db         64 CATACTGGAG 55


RESULT 10
US-09-949-016-44799
; Sequence 44799, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44799
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-44799

  Query Match             7.0%;  Score 86.2;  DB 4;  Length 601;
  Best Local Similarity   54.4%;  Pred. No. 2.1e-10;
```

```
  Matches  241;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    3;

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              | | | | | || |    | || || |||||  ||||  || || | ||    |  | |||
Db         56 CATACTAGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCATTTATATGGTCCTCA 115

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                ||    ||||   || |  |||| |   ||   ||| || ||||||      ||
Db        116 ACCCTAACTAGACATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGT------GAA 169

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              || ||||  ||| | |||  |  |   |||     |    |||||||   ||   ||| |
Db        170 GAATGTGGCAAATCTTTTAGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACT 229

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              | | | ||||| | | | |||  ||  ||||| ||||   |||| |       |    |
Db        230 GGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACTGGTCCTCAACTCTT 289

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |  | |||| | || |  |||||  |          |  | || ||||||| |||    ||
Db        290 ACTAAACATAASATAATTCATACTGAAGAGAAACCCTACAAATGTGAAAAATGTGGCAAA 349

Qy        361 GCATTTTA-TAAACATCAATCTTTA--AGACATCATATATTATCTGTTCATGAAAAAACA 417
              || ||| |  |  | ||||||| |||  |  ||| | | | |   | |   || ||| |
Db        350 GCCTTTAAGCAGTCTTCAATCCTTACTAACCATAAGAGAATTCATACTGGAGAGAAACCC 409

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              |  | ||| || |||||  ||| ||| |  ||  ||||||  ||  |  |||||||
Db        410 TATAAATGTGAAGAATGTGGCAAATCTTTTAACCGGTCTTCAACTTTTACTAAACATAAG 469

Qy        478 TTAAAACATCATGGTGGATCTCC 500
               |||  |||  ||| | |   ||
Db        470 GTAATTCATACTGGAGTAAAACC 492


RESULT 11
US-09-949-016-13027
; Sequence 13027, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13027
;   LENGTH: 45138
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(45138)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13027

  Query Match             7.0%;  Score 86.2;  DB 4;  Length 45138;
  Best Local Similarity   54.4%;  Pred. No. 8.4e-10;
  Matches  241;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    3;

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              | | | | | || |    | || || |||||  ||||  || || | ||    |  | |||
Db      42121 CATACTAGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCATTTATATGGTCCTCA 42180

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                ||    ||||   || |  |||| |   ||   ||| || ||||||      ||
Db      42181 ACCCTAACTAGACATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGT------GAA 42234

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              || ||||  ||| | |||  |  |   |||     |    |||||||   ||   ||| |
Db      42235 GAATGTGGCAAATCTTTTAGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACT 42294

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              | | | ||||| | | | |||  ||  ||||| ||||   |||| |       |    |
Db      42295 GGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACTGGTCCTCAACTCTT 42354

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |  | |||| | || |  |||||  |          |  | || ||||||| |||    ||
Db      42355 ACTAAACATAAGATAATTCATACTGAAGAGAAACCCTACAAATGTGAAAAATGTGGCAAA 42414

Qy        361 GCATTTTA-TAAACATCAATCTTTA--AGACATCATATATTATCTGTTCATGAAAAAACA 417
              || ||| |  |  | ||||||| |||  |  ||| | | | |   | |   || ||| |
Db      42415 GCCTTTAAGCAGTCTTCAATCCTTACTAACCATAAGAGAATTCATACTGGAGAGAAACCC 42474

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              |  | ||| || |||||  ||| ||| |  ||  ||||||  ||  |  |||||||
Db      42475 TATAAATGTGAAGAATGTGGCAAATCTTTTAACCGGTCTTCAACTTTTACTAAACATAAG 42534

Qy        478 TTAAAACATCATGGTGGATCTCC 500
               |||  |||  ||| | |   ||
Db      42535 GTAATTCATACTGGAGTAAAACC 42557


RESULT 12
US-09-949-016-27360/c
; Sequence 27360, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27360
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-27360

  Query Match             6.9%;  Score 85.6;  DB 4;  Length 601;
  Best Local Similarity   51.4%;  Pred. No. 2.9e-10;
  Matches  252;  Conservative    0;  Mismatches  229;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |      |  || ||    | ||      |    |  | || | |
Db        499 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 440

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||    |
Db        439 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 380

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |  ||||   | ||  || | ||| ||   ||| |   ||  ||    || || |||
Db        379 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 326

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||||| || |    |||    |      |||||    |   ||| ||  | |
Db        325 GACAAAGCATTTATCTGGTCCTCARCCCTTACTGAACATAAGAGAATTCATACCAGAGAG 266
```

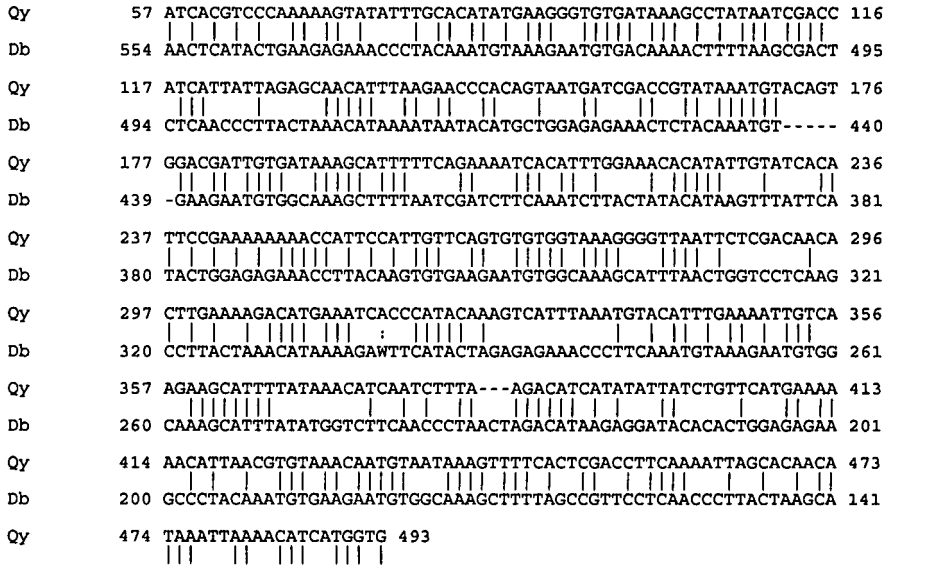


RESULT 13
US-09-949-016-149637/c
; Sequence 149637, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149637
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149637

Query Match 6.9%; Score 85.6; DB 4; Length 601;
Best Local Similarity 51.4%; Pred. No. 2.9e-10;
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

RESULT 14
US-09-949-016-27361/c
; Sequence 27361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27361
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27361

Query Match 6.7%; Score 83.2; DB 4; Length 601;
Best Local Similarity 52.5%; Pred. No. 1e-09;
Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;



```
Db        140 TAAGACAATTCATACTGGAG 121


RESULT 15
US-09-949-016-149636/c
; Sequence 149636, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149636
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-149636

  Query Match             6.7%;  Score 83.2;  DB 4;  Length 601;
  Best Local Similarity   52.5%;  Pred. No. 1e-09;
  Matches  231;  Conservative    1;  Mismatches  199;  Indels    9;  Gaps    2;

Qy         57 ATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACC 116
              | | | | |  || ||  |     | || || | |||  ||||| ||| | | ||| ||||
Db        554 AACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAAGCGACT 495

Qy        117 ATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGT 176
               |||    |     |||||  || ||  ||   |   ||   ||  || ||||||
Db        494 CTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATGT----- 440

Qy        177 GGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACA 236
               || || ||||  ||||| |||    ||   ||| || |     | |||||   |    ||
Db        439 -GAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 381

Qy        237 TTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACA 296
              | | | | |||||  | | | |||  ||  ||||| ||||   |||| |          |
Db        380 TACTGGAGAGAAACCTTACAAGTGTGAAGAATGTGGCAAAGCATTTAACTGGTCCTCAAG 321

Qy        297 CTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCA 356
              | | |  | |||| |||  :  ||||| |            |  | || | || | |||
Db        320 CCTTACTAAACATAAAAGAWTTCATACTAGAGAGAAACCCTTCAAATGTAAAGAATGTGG 261

Qy        357 AGAAGCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAA 413
                ||||||||         |  | |  ||   |||||| | |   ||      |   || ||
Db        260 CAAAGCATTTATATGGTCTTCAACCCTAACTAGACATAAGAGGATACACACTGGAGAGAA 201

Qy        414 AACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACA 473
               | |  |  ||| || |||||   |||| ||| |  ||  | ||||   |  |  | ||
Db        200 GCCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGTTCCTCAACCCTTACTAAGCA 141

Qy        474 TAAATTAAAACATCATGGTG 493
              |||   ||  |||  ||| |
Db        140 TAAGACAATTCATACTGGAG 121


Search completed: February  9, 2005, 07:51:09
Job time : 227.728 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 05:36:40 ; Search time 2154.88 Seconds
(without alignments)
3301.382 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4313806 seqs, 2877871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1232.8	99.7	1239	16	US-10-032-585-6493	Sequence 6493, Ap
2	99.4	8.0	2132	17	US-10-094-749-726	Sequence 726, App
3	96.6	7.8	2320	17	US-10-305-720-1054	Sequence 1054, Ap
c 4	92	7.4	994	16	US-10-029-386-20763	Sequence 20763, A
5	92	7.4	2873	18	US-10-755-889-369	Sequence 369, App
6	91.8	7.4	2110	17	US-10-104-047-1778	Sequence 1778, Ap
7	91.8	7.4	3078	17	US-10-104-047-622	Sequence 622, App
8	90.4	7.3	2114	17	US-10-094-749-692	Sequence 692, App
9	90.2	7.3	2230	17	US-10-094-749-499	Sequence 499, App
10	89.8	7.3	590	16	US-10-029-386-13298	Sequence 13298, A
11	89.2	7.2	2221	17	US-10-108-260A-29	Sequence 29, Appl
12	89.2	7.2	3839	17	US-10-172-118-910	Sequence 910, App
13	89.2	7.2	3839	17	US-10-342-887-910	Sequence 910, App
14	89	7.2	3502	13	US-10-098-841-48	Sequence 48, Appl
15	88.6	7.2	1466	16	US-10-029-386-25074	Sequence 25074, A
c 16	88.4	7.2	1410	16	US-10-029-386-20532	Sequence 20532, A
17	88.4	7.2	2509	17	US-10-381-327-11	Sequence 11, Appl
18	88.4	7.2	4132	14	US-10-198-846-13133	Sequence 13133, A
19	87.6	7.1	583	16	US-10-029-386-13093	Sequence 13093, A
20	87.6	7.1	1757	18	US-10-203-052B-6	Sequence 6, Appli
21	87.2	7.1	831	17	US-10-264-049-975	Sequence 975, App
22	86.8	7.0	2662	17	US-10-363-616-66	Sequence 66, Appl
23	86.2	7.0	911	16	US-10-029-386-22636	Sequence 22636, A
24	86.2	7.0	1422	16	US-10-029-386-20322	Sequence 20322, A
25	86.2	7.0	1725	16	US-10-029-386-20556	Sequence 20556, A
c 26	86.2	7.0	1727	16	US-10-029-386-20367	Sequence 20367, A
27	86.2	7.0	2558	9	US-09-764-864-1664	Sequence 1664, Ap
c 28	86.2	7.0	2558	9	US-09-764-864-1665	Sequence 1665, Ap
29	86	7.0	2664	17	US-10-108-260A-511	Sequence 511, App
30	86	7.0	2690	18	US-10-723-860-8039	Sequence 8039, Ap
31	86	7.0	3400	17	US-10-104-047-1037	Sequence 1037, Ap
32	85.4	6.9	2505	17	US-10-108-260A-808	Sequence 808, App
33	84.8	6.9	2239	17	US-10-094-749-1084	Sequence 1084, Ap
34	84.4	6.8	1191	16	US-10-029-386-22992	Sequence 22992, A
35	84.4	6.8	3138	17	US-10-104-047-83	Sequence 83, Appl
c 36	84.2	6.8	1578	16	US-10-029-386-20214	Sequence 20214, A
37	84	6.8	565	16	US-10-029-386-3970	Sequence 3970, Ap
38	84	6.8	1619	18	US-10-422-522-46	Sequence 46, Appl
39	84	6.8	3617	16	US-10-029-386-22868	Sequence 22868, A
40	84	6.8	5820	17	US-10-363-616-44	Sequence 44, Appl
c 41	83.6	6.8	596	16	US-10-029-386-6664	Sequence 6664, Ap
c 42	83.2	6.7	1101	16	US-10-029-386-25405	Sequence 25405, A
43	83.2	6.7	2241	17	US-10-108-260A-1289	Sequence 1289, Ap
44	83.2	6.7	2525	17	US-10-108-260A-1358	Sequence 1358, Ap
c 45	83	6.7	500	16	US-10-029-386-6818	Sequence 6818, Ap

ALIGNMENTS

RESULT 1
US-10-032-585-6493
; Sequence 6493, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6493
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6493

Query Match 99.7%; Score 1232.8; DB 16; Length 1239;
Best Local Similarity 99.8%; Pred. No. 9.9e-227;
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
```

```
Db       121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy       181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy       241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy       301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy       361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy       421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy       481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy       541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy       601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy       661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy       721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy       781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy       841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy       901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy       961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy      1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy      1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy      1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy      1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
             ||||||||||||||||||||||||||||||||||||
Db      1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
```

RESULT 2
US-10-094-749-726
; Sequence 726, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 726
; LENGTH: 2132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-726

Query Match 8.0%; Score 99.4; DB 17; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4e-09;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

```
Qy         7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
             ||| ||| | || |           || ||  |  |  || |    |      |   | | |
Db      1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy        67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | || |    | || || ||||| ||||  ||||| |    |||| | |||  | |
Db      1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy       127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |||||    | ||  || | |    ||    ||| || | ||||      || || |||
Db      1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy       187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
             |  ||||| |||  |       |||||  | |  ||||||| |||   ||| | | | |
Db      1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy       247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
             ||||| | ||| |||  ||  ||||||||||   ||||        |||| | |  |||
Db      1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy       307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
             ||| | |  |  |||||              |    || |||||| |||    |||| |||
Db      1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy       367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               |  |       |  | |  ||||| | | | |   |  | ||| |||   | |
Db      1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797
```

```
Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964
```

```
RESULT 3
US-10-305-720-1054
; Sequence 1054, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
;  APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
;  TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expressic
;  FILE REFERENCE: PA-0002-1 CON
;  CURRENT APPLICATION NUMBER: US/10/305,720
;  CURRENT FILING DATE:  2002-11-26
;  PRIOR APPLICATION NUMBER: 09/016,434
;  PRIOR FILING DATE: 1998-01-30
;  NUMBER OF SEQ ID NOS: 1490
;  SOFTWARE: PERL Program
; SEQ ID NO 1054
;   LENGTH: 2320
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: GenBank ID No. US20040010136A1 g1017721
US-10-305-720-1054

  Query Match             7.8%;  Score 96.6;  DB 17;  Length 2320;
  Best Local Similarity   51.3%;  Pred. No. 1.4e-08;
  Matches  306;  Conservative    0;  Mismatches  279;  Indels   12;  Gaps    3;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db       1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy        369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db       1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db       1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy        486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
Db       1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy        546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
Db       1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy        606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db       1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551
```

```
RESULT 4
US-10-029-386-20763/c
; Sequence 20763, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20763
;   LENGTH: 994
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC010620.3
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
;   OTHER INFORMATION: NT HIT: gi13650821, EVALUE 0.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: BI093435.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUE 0.00e+00
US-10-029-386-20763

  Query Match             7.4%;  Score 92;  DB 16;  Length 994;
  Best Local Similarity   52.0%;  Pred. No. 7.7e-08;
  Matches  259;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db        763 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 704

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db        703 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 644

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db        643 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 590

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db        589 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 530

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db        529 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 470

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db        469 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 410

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db        409 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 350
```

```
Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |   |   | |||||| |   |   ||||||  |||
Db        349 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 290

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |     |||
Db        289 CATACTGGAGAGAAACCT 272
```

RESULT 5
US-10-755-889-369
; Sequence 369, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 369
; LENGTH: 2873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-369

Query Match 7.4%; Score 92; DB 18; Length 2873;
Best Local Similarity 52.2%; Pred. No. 1.2e-07;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||    || ||      |     |   |
Db       1621 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 1680

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              |  | || | | || || ||||||  | ||  | ||||| ||| |  | |||     |
Db       1681 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 1740

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |  |||||   ||||  || | |    ||   ||| || |||||||  |      || |||
Db       1741 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 1794

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| ||| |  |  |||||  |    | |||||    |  ||| | ||| |
Db       1795 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 1854

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  ||||| ||||   ||||       |     | | |  | |
Db       1855 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 1914

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||   |  | ||    |  | || ||| || |||     ||| |||
Db       1915 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1974

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
               |  ||     |  | |  | ||||  |||  |  |     || |||   |
Db       1975 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 2034

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||    |||| ||| |  ||| | |||||| |   |  |||||||  |  |
Db       2035 TGTGAAGAATGCGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGTACATT 2094

Qy        484 CATCATGGTG 493
              |||  ||| |
Db       2095 CATACTGGAG 2104
```

RESULT 6
US-10-104-047-1778
; Sequence 1778, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1778
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1778

Query Match 7.4%; Score 91.8; DB 17; Length 2110;
Best Local Similarity 51.7%; Pred. No. 1.1e-07;
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

```
Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              | | || || || |    | || || | |||  ||||  |||||  | ||| |   | |||
Db       1253 CATACCGAAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAAACACTCCTCA 1312

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                  |    ||||  ||||  |||| |   ||   ||| || ||||||      ||
Db       1313 GCCCTTACTACACATAAGAGAATTCACACTGGAGAGAAACCCTACAAATGT------GAA 1366

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              || ||||  ||||| |||  | ||   ||| |  |      |||||     |   ||| |
Db       1367 GAATGTGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGAAACTTCATACT 1426

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              | ||| ||||| | | | ||| ||  |||||  ||||   |||          | | |
Db       1427 GGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTATCCAATCCTCAAAACTT 1486

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |    |||| |||  |  ||| |       |    |  |  | |||| | |||    ||
Db       1487 ACTGAACATAAAAAAATTCATTCTGGAGAGATACCCTACAAGTGTGAAGAATGTGGCAAA 1546

Qy        361 GCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAAAACA 417
              || ||| |  |        ||| |    | |||| | | |  |   |  |  || ||| |
Db       1547 GCTTTTAAACACTCCTCATCCCTTACTACACATAAAAGAATTCATACTGGGGAGAAACCC 1606

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              |  |  ||| || |||||   |||| ||| |  ||| | ||||||  |  |  |||||||
Db       1607 TACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGATCCTCAAAACTTACTGAACATAAG 1666

Qy        478 TTAAAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
               |||  |||  ||| | ||   |||     ||    | |||     |      ||| |  ||
Db       1667 ATAATTCATACTGGAGAGAAACCCTATAAATGTGAGAGATGTGACAAAGCTTTTAACCAA 1726

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
              | | ||||| |      || || |  |    |  |  |  |||| |   |||    | |
Db       1727 TCTGCAAACCT------TACTAAACATAAAAAAATACATACTGGAGAGAAACTACAGAAC 1780

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
              |   |  |||| ||| | || ||    |   |  ||| |||  | |||    || |||
Db       1781 TGGAATGTGTGATAATGATTTTGACAACACCTCAAATTTTTCTAAATATAAAGGAAATCA 1840

Qy        657 TGATGAT 663
              |  || |
Db       1841 TACTGGT 1847
```

RESULT 7
US-10-104-047-622
; Sequence 622, Application US/10104047
; Publication No. US20030236392A1

```
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
;   LENGTH: 3078
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-622

  Query Match             7.4%;  Score 91.8;  DB 17;  Length 3078;
  Best Local Similarity   52.1%;  Pred. No. 1.3e-07;
  Matches  257;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy         11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
              ||| | || |    |  ||    ||   | || ||      |     |   | | | | |
Db        650 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 709

Qy         71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
              |||||  || || || |||||  |||| |||||||| |||  |  | |||     |
Db        710 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 769

Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
              |||||   | ||  || | |   ||   ||| || ||||||     || || ||||  |
Db        770 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 823

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
              ||||  |||  | |    ||| || |    | |||||   ||   ||| | | | | ||||
Db        824 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 883

Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | | |||  ||  ||||| ||||   ||||          |  | | | |  ||||
Db        884 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 943

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  |||||| |          |  | || |||| | |||    | || ||| |
Db        944 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 1003

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
              |        | |    | |||| | | | |   | | | || ||| | |  |  |||
Db       1004 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 1063

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
              || |||||    ||||  || |     |   ||||    |   | |||||||  |||  |||
Db       1064 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 1123

Qy        488 ATGGTGGATCTCC 500
               ||| | |   ||
Db       1124 CTGGGGAAAAACC 1136


RESULT 8
US-10-094-749-692
; Sequence 692, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 692
;   LENGTH: 2114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-692

  Query Match             7.3%;  Score 90.4;  DB 17;  Length 2114;
  Best Local Similarity   51.8%;  Pred. No. 2.1e-07;
  Matches  258;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |      || ||      |  || |   |      |   | | |
Db       1024 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1083

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | || |    | || || |||||  |||| |||||| | ||| |  || |||    |
Db       1084 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTAACCACCCTTCAGCCCTT 1143

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||      |  ||   ||| ||   ||| || ||||||      || || |||
Db       1144 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1197

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||   | ||   ||| |  |    | |||||   ||   ||||| | | |
Db       1198 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1257

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||   |   ||||| |||||  ||||  |      |  | | | || |
Db       1258 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1317

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  |  |||||                |  | || |||| | |||    |||| |||
Db       1318 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1377

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||     |   ||   |  | |||| | ||  |   |  |   || ||| | |  |
Db       1378 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1437

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |  |   | ||||||  |  |   ||||||  |||
Db       1438 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1497

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |     |||
Db       1498 CATACTGGAGAGAAACCT 1515


RESULT 9
US-10-094-749-499
; Sequence 499, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
```

```
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 499
;   LENGTH: 2230
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-499

  Query Match             7.3%;  Score 90.2;  DB 17;  Length 2230;
  Best Local Similarity   50.4%;  Pred. No. 2.4e-07;
  Matches  332;  Conservative    0;  Mismatches  313;  Indels   14;  Gaps    4;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||||| || |    |  ||    ||     || ||     |     |   | | |
Db        656 GAATGTGACAAATCACTTTGCATGCTTTCACGCCTAACTCAACATAAAAAAATTCATACT 715

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | | || || | | || || ||||| |||||  ||| ||| |||  |  |  ||    |
Db        716 AGAGAGAATTTCTACAAATGTGAAGAGTGTGGAAAAACCTTTAACTGGTCCACAAACCTT 775

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||  |   | ||  || | |   ||   ||| || ||||||  |||       |||
Db        776 TCTAAACCTAAGAAAATTCATACTGGAGAAAAACCCTACAAATGTGAAGTA------TGT 829

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |  |   |||     |    | |||||   ||   | | | | | ||
Db        830 GGAAAAGCCTTTCACCAATCCTCAATCCTTACTAAACATAAGATAATTCGTACTGGAGAA 889

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  |   | ||| ||   ||||| ||||   ||||          |||| | |  |||
Db        890 AAACCCTATAAATGTGCACACTGTGGCAAAGCCTTTAAACAGTCCTCACACCTTACTAGA 949

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  | || |  |||||  |          |  | || |||| | |||     | |  |||
Db        950 CATAAGATAATTCATACTGAAGAGAAACCCTACAAATGTGAACAATGTGGCAAGGTCTTT 1009

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |    | | |   |  | |||||| ||| | | |  |   ||  || ||| |
Db       1010 AAGCAGTCCCCAACCCTTACTAAACATCAGATAATTTATACTGGAGAGGAACCATACAAA 1069

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| |  |||||   |||| ||| |  | | ||| | || |  |  |||||||   ||
Db       1070 TGTGAGGAATGTGGCAAAGCTTTTAACCTATCTTAACAACTTACTGAACATAAGAAAATT 1129

Qy        484 CATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCA 542
              |   | | ||   ||| |  ||       ||     |      ||| |  | ||| ||
Db       1130 TACACTAGAGAGAAAGCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTTTTCA 1189

Qy        543 AACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAA 602
              | | |  | |    ||  ||     |||  | |     | ||    |  ||||||    |
Db       1190 ACCCTTATTACACATAAGATAATTCATAGCAGAGAGAAACCC----CACAAATGTGAAGA 1245

Qy        603 ATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATG 661
              |||||| | || || |     |     |    |  ||  ||||| ||||| ||||  ||
Db       1246 ATGTGGCAGAGCTTTTAACCAGTCCGCAAAGCTCACTGAACATAAGTTAATTCATACTG 1304


RESULT 10
US-10-029-386-13298
; Sequence 13298, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR (
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13298
;   LENGTH: 590
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR19.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
;   OTHER INFORMATION: NT HIT: gi14755315, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALUE 1.00e-119
;   OTHER INFORMATION: EST_HUMAN HIT: BG619358.1, EVALUE 0.00e+00
US-10-029-386-13298

  Query Match             7.3%;  Score 89.8;  DB 16;  Length 590;
  Best Local Similarity   53.8%;  Pred. No. 1.7e-07;
  Matches  234;  Conservative    0;  Mismatches  192;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
              || | ||| || | || | ||  | |||||  |||| |||||  | |    |  | |||
Db        126 GTGCTAAAGAGAAAATCTACAAGTGTGAAGAATGTGGCAAAGCATTTCTATGGTCCTCAA 185

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
              ||     ||||   || |  |||| |   ||  ||| || ||||||     || |
Db        186 CCCTAACTAGACATAAGAGGATACACACTGGAGAGAAACCCTACAAATGT------GAAG 239

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
              | ||||  ||||| |||  |       |||     | |  | |||||     |   ||| | |
Db        240 AATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTG 299

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
              | | ||||| | | | |||  ||  ||||| ||||   |||     |     | | |
Db        300 GAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGTTCTTCAACCCTTG 359

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
              | |||| | |  |  ||||||          | | || | || | |||     |||
Db        360 CTAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAG 419

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAA---AAAACAT 418
              | |||   ||      | |  |      ||||||   || ||  |  ||||   ||| | |
Db        420 CTTTTAGCAATTCCTCAACCCTTGCTAATCATAAGATAACTCATACTGAAGAGAAACCCT 479

Qy        419 TAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAAT 478
              |  |||||| ||||| | |||| ||| |  ||||  ||||    |  |  ||||||||
Db        480 ACAAATGTAAAGAATGTGACAAAGCTTTTAAGCGACTCTCAACCCTTACTAAACATAAAA 539

Qy        479 TAAAACATCATGGTG 493
              ||| ||||  ||| |
Db        540 TAATACATGCTGGAG 554


RESULT 11
US-10-108-260A-29
```

```
; Sequence 29, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
;   LENGTH: 2221
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-29

  Query Match             7.2%;  Score 89.2;  DB 17;  Length 2221;
  Best Local Similarity   53.4%;  Pred. No. 3.7e-07;
  Matches  237;  Conservative    0;  Mismatches  198;  Indels    9;  Gaps    2;

Qy          61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
               | | ||  | || |   || || || | | |  |||| ||||||| |||| ||   |||
Db         654 CATACCGGAGAGAAACCTTTCAGATGTAAGGAATGTGGCAAAGCCTTTAATAGAAGCTCA 713

Qy         121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
               |    |     || |||    ||||  || | |    ||   ||| |  || ||||       |
Db         714 TACGTTACTCAGCATCAGAGAATGCATACTGGAGAGAAACCTTTCAAGTGTA------AG 767

Qy         181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
               || ||||  ||||| |||  ||||   |||||  |      ||||     |   ||| |
Db         768 GAATGTGGCAAAGCTTTTAACAGAGCTTCACACCTTACTCAACATCAGAGAATCCATACT 827

Qy         241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
               | | | ||||  ||| | |||  ||  ||||| ||||   ||||      |   |||| |
Db         828 GGAGAGAAACACTTCAAATGTAAAGAATGTGGCAAAGCCTTTAACAGGGGCTCACACCTT 887

Qy         301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
               |    ||||| |||  | ||||||                | | || |||| | |||     ||
Db         888 ACTCGACATCAAAGAATCCATACTGGAGAGAAGTCTTTCAAATGTGAAGAATGTGGCAAA 947

Qy         361 GCATTT---TATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACA 417
               || |||   |      ||   |    ||||| | | |   |  |   || |||
Db         948 GCTTTTATCTGGGGCTCACACCTTACTCAACATCAGAGAGTCCATACTGGAGAGAAATTC 1007

Qy         418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
               || |  |||||| |||||   |||| ||| ||  |    ||| |  |  | ||||||
Db        1008 TTCAAATGTAAAGAATGTGGCAAAGCTTTTACCAGGAGCTCACACCTTACTCAACATCAG 1067

Qy         478 TTAAAACATCATGGTGGATCTCCT 501
                  ||  |||  ||| |     |||
Db        1068 AGAATTCATACTGGAGAGAAACCT 1091


RESULT 12
US-10-172-118-910
; Sequence 910, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Chris
;  APPLICANT:  Van 't Veer, Laura
;  APPLICANT:  Van de Vijver, Marc
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-175-999
;  CURRENT APPLICATION NUMBER: US/10/172,118
;  CURRENT FILING DATE:  2002-06-14
;  PRIOR APPLICATION NUMBER: 60/380,770
;  PRIOR FILING DATE: 2002-05-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
;   LENGTH: 3839
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: NM_003430
;   DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-910

  Query Match             7.2%;  Score 89.2;  DB 17;  Length 3839;
  Best Local Similarity   53.7%;  Pred. No. 4.6e-07;
  Matches  234;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    2;

Qy          69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
               | || |    | || || |||||  ||||  ||||| | ||  |   | |||    | |
Db         964 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1023

Qy         129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                 |||||   ||||  || | |   ||   ||| || ||||||       || || ||||
Db        1024 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1077

Qy         189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               |||||  |||  |        |||     | |  | |||||     |   ||| | | | | ||
Db        1078 CAAAGCTTTTAGCCATTCTTCAGCCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 1137

Qy         249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
               ||| | | | |||  ||  ||||| ||||   |||    |      |  | |    |  ||
Db        1138 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 1197

Qy         309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
               | | || || |||||  |           |  | || | || | ||| |  || | ||| |
Db        1198 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 1257

Qy         369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                 ||     | |   |   | |||| | ||| ||  || |   || |||    |  | ||
Db        1258 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1317

Qy         426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
               | || |||||   |||| ||| | |||| |||||||  |  |   |||||| || |  ||
Db        1318 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1377

Qy         486 TCATGGTGGATCTCCT 501
               |  ||| |     |||
Db        1378 TACTGGAGAGAAACCT 1393


RESULT 13
US-10-342-887-910
; Sequence 910, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
;   LENGTH: 3839
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-10-342-887-910

  Query Match             7.2%;  Score 89.2;  DB 17;  Length 3839;
  Best Local Similarity   53.7%;  Pred. No. 4.6e-07;
  Matches  234;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||   |   | |||    | |
Db        964 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  |      |||    | |  | |||||    |   |||  | | | | ||
Db       1078 CAAAGCTTTTAGCCATTCTTCAGCCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | | | |||  ||  ||||| ||||   |||    |      |  | |    |  ||
Db       1138 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || ||||  |           | | || | || | |||  |  || | |||  |
Db       1198 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 1257

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
               ||    | |   |  |||| | ||| ||  || |   || |||   |  |  ||
Db       1258 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||   |  |   ||||| || |  ||
Db       1318 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1377

Qy        486 TCATGGTGGATCTCCT 501
              |  |||  |     |||
Db       1378 TACTGGAGAGAAACCT 1393


RESULT 14
US-10-098-841-48
; Sequence 48, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Xu, Chongjun
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Qian, Xiaohong B.
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2
;  CURRENT APPLICATION NUMBER: US/10/098,841
;  CURRENT FILING DATE:  2002-03-13
;  PRIOR APPLICATION NUMBER: 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 331
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 48
;   LENGTH: 3502
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (154)..(2583)
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(3502)
;   OTHER INFORMATION: n = a,t,c or g
US-10-098-841-48

  Query Match             7.2%;  Score 89;  DB 13;  Length 3502;
  Best Local Similarity   51.7%;  Pred. No. 4.8e-07;
  Matches  256;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy         79 TGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTA 138
              | || || |||||  ||||  ||||||| |||  |  | |||     |    |||||
Db       2101 TACAAATGTGAAGAATGTGGCAAAGCCTTTAAGTGGTCCTCAACCCTTACTAAACATAAG 2160

Qy        139 AGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTT 198
              | ||  || | |   ||   ||| || ||||||      || || ||||  ||||| |||
Db       2161 ATAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGGCAAAGCTTTT 2214

Qy        199 TTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCAT 258
                  |||    |   |||||||   |    ||| | | | | ||||| | | |
Db       2215 AAACTGTCCTCAACCCTTTCTACACATAAGATTATTCATACTGGAGAGAAACCCTACAAA 2274

Qy        259 TGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACC 318
              |||  |   ||||| ||||   ||||     |    | || | |    |||| | | |
Db       2275 TGTGAAAAATGTGGCAAAGCTTTTAACCGACCCTCAAACCTTATTGAACATAAGAAAATT 2334

Qy        319 CATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT---TATAAACAT 375
              |||||              |  | || |||| | |||     ||||||||   |||
Db       2335 CATACTGGAGAGCAACCCTACAAATGTGAAGAATGTGGCAAAGCATTTAACTATTCCTCA 2394

Qy        376 CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGT 435
              ||   |   | |||| | | | |   |  | | ||  || | |  |  ||||| |||||
Db       2395 CACCTTAATACACATAAGAGAATTCATACTAAAGAGCAACCCTACAAATGTAAAGAATGT 2454

Qy        436 AATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGA 495
                 |||| |||||  | |  ||||||  |  |   ||||||   ||  |||  ||| |
Db       2455 GGCAAAGCTTTCAACCAATATTCAAACCTTACTACACATAACAAAATTCATACTGGAGAG 2514

Qy        496 TCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTA 555
                 |    | |     |||    | |    |||  | ||     |||||||  |||
Db       2515 AAACTCTACAAACCTGAAGATGTGACAGTGATTTTGACAACACCTCAAACTTTTTCAAAC 2574

Qy        556 TTACAATTTCATATA 570
               || |||    ||||
Db       2575 ATAAAATAAATTATA 2589


RESULT 15
US-10-029-386-25074
; Sequence 25074, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25074
;   LENGTH: 1466
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC011467.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
;   OTHER INFORMATION: NT HIT: gi15309152, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P35789, EVALUE 0.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AU143734.1, EVALUE 0.00e+00
US-10-029-386-25074

  Query Match             7.2%;  Score 88.6;  DB 16;  Length 1466;
  Best Local Similarity   53.3%;  Pred. No. 4.1e-07;
  Matches  237;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |   |||| |   || || || | |||  ||||| ||  | | |       |  ||
Db        272 ACATACTGGAAAGAAATCTTTCAAATGTAAAGAATGTGAAAAGTCATTTTGCATGCTTTC 331

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
              |   ||||  ||||||  |||||  || |||   ||   ||| || |||||||
Db        332 ACACTTAGCTCAACATAAAAGAATTCATAGTGGAGAGAAACCCTACAAATGTAAA----- 386

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
               || ||||  ||||| | |       |  ||| |  |    |||||||    |   ||| |
Db        387 -GAATGTGGGAAAGCCTATAATGAGACCTCAAACCTTTCTACACATAAAAGAATTCATAC 445

Qy        240 CGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
              | ||| ||||| | | | ||   || |||||| ||||   ||||    |    |||| |
Db        446 TGGAAAGAAACCCTACAAATGCGAAGAGTGTGGAAAAGCCTTTAACCGGCTCTCACACCT 505

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
              |  | |||| | || |  |||||             |  | || |||  | |||    |
Db        506 TACTACACATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAGGAGTGTGGCAA 565

Qy        360 AGCATTTTATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
              ||| ||| |  ||  |    |   |   | |||| | | | |   |  |   || ||| |
Db        566 AGCTTTTAACCAATCTGCAAACCTTACTACACATAAGAGAATTCATACTGGAGAGAAACC 625

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
              |  |  ||| || |||||   | || ||| |  |   | ||||   |  |   ||||||
Db        626 CTACAAATGTGAAGAATGTGGCAGAGCTTTTAGCCAGTCCTCAACCCTTACTGCACATAA 685

Qy        477 ATTAAAACATCATGGTGGATCTCCT 501
                 |||  |||  ||| |    |||
Db        686 GATAATTCATGCTGGAGAGAAACCT 710


Search completed: February  9, 2005, 11:50:30
Job time : 2157.88 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:45:39 ; Search time 3806.38 Seconds
(without alignments)
12360.150 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	182.2	14.7	628	8	BZ298000	BZ298000 CG3957.f1
c	2	167.6	13.6	959	9	CNS06MMP	AL405671 T7 end of
	3	103.8	8.4	801	7	CN759437	CN759437 ID0AAA25C
	4	103.8	8.4	898	7	CN761003	CN761003 ID0AAA2AE
	5	98.2	7.9	1609	3	AK032220	AK032220 Mus muscu
	6	97.8	7.9	2647	3	BC047646	BC047646 Homo sapi
	7	97.6	7.9	745	1	AU123448	AU123448 AU123448
	8	93.4	7.6	727	7	CO434402	CO434402 UI-M-HX0-
	9	92.4	7.5	631	7	CN352904	CN352904 170006000
	10	92.4	7.5	2672	3	BC036394	BC036394 Homo sapi
	11	92	7.4	591	8	AQ005136	AQ005136 CIT-HSP-2
	12	92	7.4	687	7	CN362485	CN362485 170004241
	13	92	7.4	752	7	CN417582	CN417582 170004240
	14	91	7.4	593	8	B99387	B99387 CIT-HSP-228
	15	91	7.4	820	5	BX414627	BX414627 BX414627
	16	90.8	7.3	666	7	CN297781	CN297781 170005321
	17	90.4	7.3	2694	3	BC022527	BC022527 Homo sapi
	18	90.4	7.3	2698	3	BC037782	BC037782 Homo sapi
	19	90.2	7.3	2253	3	BC028252	BC028252 Mus muscu
	20	89.8	7.3	550	5	BP292533	BP292533 BP292533
	21	89.2	7.2	582	5	BP234906	BP234906 BP234906
	22	89.2	7.2	604	2	BE161630	BE161630 MR3-HT044
	23	89.2	7.2	638	7	CN341035	CN341035 170006000
	24	89.2	7.2	673	7	CN353930	CN353930 170006001
	25	89.2	7.2	3609	3	BC032590	BC032590 Homo sapi
	26	89	7.2	2856	3	HSM803730	AL832422 Homo sapi
	27	88.6	7.2	535	1	AL705393	AL705393 DKFZp686M
	28	88.6	7.2	612	5	BP238960	BP238960 BP238960
	29	88.6	7.2	2174	3	AK033001	AK033001 Mus muscu
	30	88.6	7.2	2330	3	BC037426	BC037426 Homo sapi
	31	88.6	7.2	2622	3	AK033958	AK033958 Mus muscu
	32	88.4	7.2	998	5	BX456765	BX456765 BX456765
	33	88	7.1	895	5	BU195416	BU195416 AGENCOURT
	34	87.8	7.1	608	7	CN266758	CN266758 170004554
	35	87.8	7.1	4251	3	CR749856	CR749856 Homo sapi
	36	87.6	7.1	744	5	BX437291	BX437291 BX437291
	37	87.2	7.1	579	8	AQ347265	AQ347265 RPCI11-11
c	38	87.2	7.1	687	8	AQ389266	AQ389266 RPCI11-15
	39	87.2	7.1	907	5	BQ423752	BQ423752 AGENCOURT
	40	87	7.0	660	6	CD770140	CD770140 AGENCOURT
	41	87	7.0	842	5	BU507408	BU507408 AGENCOURT
c	42	86.8	7.0	521	8	AQ475626	AQ475626 CITBI-E1-
	43	86.8	7.0	571	7	CN280110	CN280110 170006001
	44	86.8	7.0	705	8	AQ194282	AQ194282 RPCI11-61
c	45	86.8	7.0	860	8	AQ749175	AQ749175 HS_5575_A

ALIGNMENTS

RESULT 1
BZ298000/c
LOCUS BZ298000 628 bp DNA linear GSS 31-OCT-2002
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata genomic clone CG3957, genomic survey sequence.
ACCESSION BZ298000
VERSION BZ298000.1 GI:24440936
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 628)
AUTHORS Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), R10 (2003)
MEDLINE 22508158
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.
FEATURES Location/Qualifiers
source 1. .628
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG3957"
/clone_lib="Candida glabrata Random Genomic Library"
ORIGIN

Query Match 14.7%; Score 182.2; DB 8; Length 628;
Best Local Similarity 59.5%; Pred. No. 7.3e-26;
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

```
Qy          84 ATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAAC 143
               ||| ||    ||||| || |||| ||   | || || ||  |     ||||    |  ||
Db         625 ATACGATAATTGTGACAAGGCCTTTACAAGGCCGTCCTTGCTCACAGAACACCAGAATAC 566

Qy         144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAG 203
                  ||  | |   | |    || | |    |  | ||||| |   |||||   |
```

```
Db        565 ---AGTGCATCTTGGGAGGAAACCTTGGAAATGTAATCAATGTGAAAGTTCATTTACTAA 509

Qy        204 AAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTC 263

Db        508 AAAGATCCACTTAGAGAGACACTTATACACACACACAGATGAAAGACCGTTTTATTGTTC 449

Qy        264 AGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATAC 323

Db        448 ATTTTGTGGGAAGGGGCTAATTACTAGGCAACAACTGAAACGACATGAGGTAACTCACAC 389

Qy        324 AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTT 383

Db        388 CAAATCGTTCAATTGTGAATATGAAGGTTGTAATGAGAGTTTCTACAAGCACCCTCAATT 329

Qy        384 AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGT 443

Db        328 AAGGGCTCATATTTTAGCAGTTCATTTACAAAGTCTAAAATGTCACGAATGCAACAAAAG 269

Qy        444 TTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCAT------GGTGGATC 497

Db        268 TTTTCAAAGACCTTACAGACTCAAAAATCACATAGCCAAACACCATAATCCAGATGTTGT 209

Qy        498 TCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATT 557

Db        208 CAATGCTTATCAATGTACCTTCAGTGTTTGTTCTAAGAGTTTCAAAACATGGTCTGCGTT 149

Qy        558 ACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTG 617

Db        148 AAGATTACATGTTAAAAATGATCATCCGAAGTTAAAATGCCCTATATGCAGTAAACCCTG 89

Qy        618 TGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662

Db         88 TGTTGGGGAAGACGGTTTAAATATGCACATGAAAATCCACGATGA 44
```

RESULT 2
CNS06MMP/c

```
LOCUS       CNS06MMP                 959 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AU0AA005F10 of library AU0AA from strain CBS 3082
            of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION   AL405671
VERSION     AL405671.1  GI:12168715
KEYWORDS    GSS.
SOURCE      Saccharomyces kluyveri
  ORGANISM  Saccharomyces kluyveri
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 959)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 959)
  AUTHORS   Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
            Gaillardin,C. and Casaregola,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 9.
            Saccharomyces kluyveri
  JOURNAL   FEBS Lett. 487 (1), 56-60 (2000)
  MEDLINE   20584719
   PUBMED   11152884
REFERENCE   3  (bases 1 to 959)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .959
                     /organism="Saccharomyces kluyveri"
                     /mol_type="genomic DNA"
                     /strain="CBS 3082"
                     /db_xref="taxon:4934"
                     /clone="AU0AA005F10"
                     /clone_lib="AU0AA"
                     /note="end : T7"
     misc_feature    complement(<71. .>958)
                     /note="similar to Saccharomyces cerevisiae ORF YPR186c [
                     PZF1 ; TFIIIA (transcription initiation factor) ]"
                     /evidence=not_experimental
ORIGIN
```

```
  Query Match             13.6%;  Score 167.6;  DB 9;  Length 959;
  Best Local Similarity   57.2%;  Pred. No. 5.9e-23;
  Matches  324;  Conservative    1;  Mismatches  235;  Indels    6;  Gaps    1;

Qy        211 CATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGT 270

Db        958 CACTTGGAAAGACATATGTTCACGCATTCCGATAATAAACCATTTCATTGTYCCTATTGT 899

Qy        271 GGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCA 330

Db        898 GGTAAAGGTGTTACTACAAAGCAGCAGTTAAAAAGACATGAGATCACACATACCAAATCG 839

Qy        331 TTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACAT 390

Db        838 TTCAAATGTGAATATGAGGGCTGCAACGAATGTTTTTACAAGCATTCTCAACTAAGGTCA 779

Qy        391 CATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACT 450

Db        778 CATACACTTTCGGTACACTTGCAGAAATTGACCTGTGAGCATTGCGGTAAAATGTTCCAA 719

Qy        451 CGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCT------ 504

Db        718 AGGCCATATCGTCTGCAAAACCATCTTGCAAAGCATCACAATGTGGATGTTGAAAACAAA 659

Qy        505 TATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTT 564

Db        658 TATCAATGTACTTACACGGCGTGCACAGAGACTTTTAAAACCTGGACTGCTCTTCAGCAG 599

Qy        565 CATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGG 624

Db        598 CATATCAAGGCGGACCATCCCAAGTTGCACTGTAACGTATGTGGCAAAGCTTGTGTTGGT 539

Qy        625 AAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGG 684

Db        538 GAGTCTGGTCTACAGATGCACATGCGAGTCCATGACGAGGCTTTAGTAATAAAAAATTGG 479

Qy        685 ACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTAT 744

Db        478 AAATGTACCATTTGTCAGCAAGTCTCTTTTGCTAAGAAGGCTGATCTTTTATCTCACTAC 419

Qy        745 AATATCTTCCATGATGGTAATATCCC 770

Db        418 ATGGAACATCATAAGGATGATATACC 393
```

RESULT 3
CN759437

```
LOCUS       CN759437                 801 bp    mRNA    linear   EST 20-MAY-2004
```

```
DEFINITION  ID0AAA25CC01RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA25CC01
            5', mRNA sequence.
ACCESSION   CN759437
VERSION     CN759437.1  GI:47533360
KEYWORDS    EST.
SOURCE      Acyrthosiphon pisum (pea aphid)
  ORGANISM  Acyrthosiphon pisum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
            Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,
            Stern,D., Tagu,D. and Wincker,P.
  TITLE     An expressed sequence tags database for the pea aphid Acyrthosiphon
            pisum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: D. Tagu
            INRA Rennes
            UMR BiO3P, BP 35327, F-35653 Le Rheu Cedex France
            Tel: +33.2.23.48.51.65
            Fax: +33.2.23.48.51.50
            Risk of contamination by bacterial sequences from obligatory
            (Buchnera) or facultative endosymbionts. These sequences were
            obtained in the frame of the International Consortium of Aphid
            Genomics in collaboration with Genoscope
            PCR PRimers
            FORWARD: CAGGAAACAGCTATGACC
            Plate: 25  row: C  column: 1.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Acyrthosiphon pisum"
                     /mol_type="mRNA"
                     /cultivar="developmentstage"
                     /db_xref="taxon:7029"
                     /clone="ID0AAA25CC01"
                     /tissue_type="whole insect"
                     /dev_stage="nymphs and adults (parthenogenetic females)"
                     /lab_host="XL1-Blue"
                     /clone_lib="ApMS"
                     /note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI;
                     Sample name: ID0AAA ; Plant growth place: Department of
                     Ecology & Evolutionary Biology, Princeton University ;
                     Soil conditions: Soil ; Sowing date: 01/06/1999 ;
                     Harvesting date: 01/06/1999 ; Stress date: no stress ;
                     Description: Aphids inoculated on one-week old Vicia faba
                     under non-sterile conditions.  All parthenogenetic stages
                     and both winged and wingless adults were collected for
                     library construction. ; experimental condition: long
                     photoperiod (16-hr light/8-hr dark at 18 c)"
ORIGIN

  Query Match             8.4%;  Score 103.8;  DB 7;  Length 801;
  Best Local Similarity   51.5%;  Pred. No. 2.6e-10;
  Matches  318;  Conservative    0;  Mismatches  287;  Indels   12;  Gaps    3;

Qy         82 ACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGA 141
              | || |||    |||||||||| | || |  |  ||||   |||      |||  ||
Db          8 AAATGTGATTTCTGTGATAAAGAGTTTACTATAATATCACGTTTAACAAGGCATACAAAG 67

Qy        142 ACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTC 201
              || || |    |||    |||||| |||||      ||  | |||||| |||| ||||
Db         68 ACACATACCGGTGAAAAGCCGTATGAATGT------GATAACTGTGATCAAGCGTTTTCT 121

Qy        202 AGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGT 261
                |||||||||||    |   ||||   |  ||||| |||   |||| || |   | |||
Db        122 TTGAAATCACATTTAATAGGCCATAAAATGACACATACCGGCGAAAAGCCTTATAAATGT 181

Qy        262 TCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCAT 321
                  |||| | ||| | ||  |     |  | | || | ||| |||  ||  || |||
Db        182 GATAACTGTGATCAAGCGTTTTCTAAGAAATCAAATTTAACAAGGCATATAAGGACACAT 241

Qy        322 ACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCT 381
              ||      |       | || || |||||| ||| |  |||| |||| | |  |   |  |

Db        242 ACCGGCGAAAAGCCGTATAAATGTGAAAAGTGTAACCAAGCGTTTTCTCAGAAATCAGAT 301

Qy        382 TTAAGACATCATATATTATCTGTTCAT---GAAAAAACATTAACGTGTAAACAATGTAAT 438
              |||| || |||||||    |   |      |||||  | |  |  ||| |  | ||| ||
Db        302 TTAATACGTCATATAAGGACACATACCGGCGAAAAGCCGTATAAATGTGATAACTGTGAT 361

Qy        439 AAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCT 498
               |||  ||  ||| |   |  |  |||       |||||||| | ||||   |||| |
Db        362 CAAGCGTTTTCTCAAAAATGTAGTTTAATTAGCCATAAATTGACACATACCGGTGAAAAG 421

Qy        499 CCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTA 558
              |   | | | |||||| | |   || |          ||    ||      | |  |
Db        422 C---CGTTTAAATGTGCTAACTGTGATCAAGCGTTTTATCGGAAAGACCATTTAAAAAGC 478

Qy        559 CAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGT 618
              || |      | | |   |   ||        |||||||  || ||||    |||  | |
Db        479 CACTCAAGGACACATACAGGCCAAAAGCCGTATAAATGTGATACATGTAACCAAGCGTTT 538

Qy        619 GTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAA 678
                |  ||||  |  |||| |    ||||     |  ||||  | |   |   || | |||
Db        539 TCTCAGAAATCAAATTTAACAAGCCATACAAGGACACATGCCGGTGAAAAGCTGTTTAAA 598

Qy        679 ATATGGACTTGTGATTA 695
                  |  |||||| |
Db        599 TGCGATAACTGTGATAA 615


RESULT 4
CN761003
LOCUS       CN761003                 898 bp    mRNA    linear   EST 20-MAY-2004
DEFINITION  ID0AAA2AE05RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA2AE05 5',
            mRNA sequence.
ACCESSION   CN761003
VERSION     CN761003.1  GI:47534926
KEYWORDS    EST.
SOURCE      Acyrthosiphon pisum (pea aphid)
  ORGANISM  Acyrthosiphon pisum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
            Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,
            Stern,D., Tagu,D. and Wincker,P.
  TITLE     An expressed sequence tags database for the pea aphid Acyrthosiphon
            pisum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: D. Tagu
            INRA Rennes
            UMR BiO3P, BP 35327, F-35653 Le Rheu Cedex France
            Tel: +33.2.23.48.51.65
            Fax: +33.2.23.48.51.50
            Risk of contamination by bacterial sequences from obligatory
            (Buchnera) or facultative endosymbionts. These sequences were
            obtained in the frame of the International Consortium of Aphid
            Genomics in collaboration with Genoscope
            PCR PRimers
            FORWARD: CAGGAAACAGCTATGACC
            Plate: 2  row: E  column: 5.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Acyrthosiphon pisum"
                     /mol_type="mRNA"
                     /cultivar="developmentstage"
                     /db_xref="taxon:7029"
                     /clone="ID0AAA2AE05"
                     /tissue_type="whole insect"
                     /dev_stage="nymphs and adults (parthenogenetic females)"
                     /lab_host="XL1-Blue"
                     /clone_lib="ApMS"
                     /note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI;
                     Sample name: ID0AAA ; Plant growth place: Department of
                     Ecology & Evolutionary Biology, Princeton University ;
                     Soil conditions: Soil ; Sowing date: 01/06/1999 ;
```

Harvesting date: 01/06/1999 ; Stress date: no stress ;
Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 898;
Best Local Similarity 51.5%; Pred. No. 2.6e-10;
Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

```
Qy         82 ACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGA 141
Db          7 AAATGTGATTTCTGTGATAAAGAGTTTACTATAATATCACGTTTAACAAGGCATACAAAG 66

Qy        142 ACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTC 201
Db         67 ACACATACCGGTGAAAAGCCGTATGAATGT------GATAACTGTGATCAAGCGTTTTCT 120

Qy        202 AGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGT 261
Db        121 TTGAAATCACATTTAATAGGCCATAAAATGACACATACCGGCGAAAAGCCTTATAAATGT 180

Qy        262 TCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCAT 321
Db        181 GATAACTGTGATCAAGCGTTTTCTAAGAAATCAAATTTAACAAGGCATATAAGGACACAT 240

Qy        322 ACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCT 381
Db        241 ACCGGCGAAAAGCCGTATAAATGTGAAAAGTGTAACCAAGCGTTTTCTCAGAAATCAGAT 300

Qy        382 TTAAGACATCATATATTATCTGTTCAT---GAAAAAACATTAACGTGTAAACAATGTAAT 438
Db        301 TTAATACGTCATATAAGGACACATACCGGCGAAAAGCCGTATAAATGTGATAACTGTGAT 360

Qy        439 AAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCT 498
Db        361 CAAGCGTTTTCTCAAAAATGTAGTTTAATTAGCCATAAATTGACACATACCGGTGAAAAG 420

Qy        499 CCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTA 558
Db        421 C---CGTTTAAATGTGCTAACTGTGATCAAGCGTTTTATCGGAAAGACCATTTAAAAAGC 477

Qy        559 CAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGT 618
Db        478 CACTCAAGGACACATACAGGCCAAAAGCCGTATAAATGTGATACATGTAACCAAGCGTTT 537

Qy        619 GTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAA 678
Db        538 TCTCAGAAATCAAATTTAACAAGCCATACAAGGACACATGCCGGTGAAAAGCTGTTTAAA 597

Qy        679 ATATGGACTTGTGATTA 695
Db        598 TGCGATAACTGTGATAA 614
```

RESULT 5
AK032220
LOCUS AK032220 1609 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430503O15 product:MSZF33 (FRAGMENT) homolog [Mus musculus], full insert sequence.
ACCESSION AK032220
VERSION AK032220.1 GI:26328050
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1609)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6430503O15"
/db_xref="taxon:10090"
/clone="6430503O15"
/sex="male"

```
                     /tissue_type="olfactory brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             330. .1223
                     /note="unnamed protein product; MSZF33 (FRAGMENT) homolog
                     [Mus musculus] (SPTR|O88251, evidence: FASTY, 87.1%ID,
                     100%length, match=255)
                     putative"
                     /codon_start=1
                     /protein_id="BAC27766.1"
                     /db_xref="GI:26328051"
                     /translation="MLETYRNLNAVGYNWEDSNIEEHCESSRRHGRHERNHTGEKPYE
                     GIQYGEAFVHHSSLQMRKIIHTGEKRYKCNQCDKAYSRHSILQIHKRTHSGEKPYECN
                     QCGKAFTQHSHLKIHMVTHTGEKPYKCNQCGKAFACHNKLQKHERTHTGEKPYKCDQC
                     NKAFVYESYLQVHKKTHTGEKPYKCNECGKAFARHSHLKVHKITHTGEKPYKCNQCGK
                     ALAYHSTLQVHQRTHTGEKPYECEQCGKAFANQSYFQVHKRIHTGEKPYKCDQCGKAF
                     VGSSDLKRHERVHTGRETLQM"
     polyA_signal    1594. .1599
                     /note="putative"
     polyA_site      1609
                     /note="putative"
ORIGIN

  Query Match             7.9%;  Score 98.2;  DB 3;  Length 1609;
  Best Local Similarity   51.2%;  Pred. No. 3.6e-09;
  Matches  310;  Conservative    0;  Mismatches  283;  Indels   12;  Gaps    3;

Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |   | || |    | || ||   |    ||||||||||||||||   ||||
Db        521 ACATACTGGAGAGAAACGCTACAAATGCAATCAATGTGATAAAGCCTATTCACGACACAG 580

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
                | || |   ||||  |||||| || |||   ||   ||| ||| ||||||      |
Db        581 CATTCTACAAATACATAAAAGAACACATAGTGGAGAGAAACCCTATGAATGTA------A 634

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
                | |||| |||||| |||     | |    ||| |  ||| |||||| ||  ||||| |
Db        635 TCAATGTGGTAAAGCCTTTACACAACACAGTCATCTCAAAATACATATGGTTACACATAC 694

Qy        240 CGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
               | | | ||||| | | | |||       ||||||||||   ||   | ||   |  | |
Db        695 TGGAGAGAAACCTTACAAATGTAATCAATGTGGTAAAGCCTTTGCATGTCATAATAAACT 754

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
                 ||| ||||||||  || |||||           |  |  | |||  | ||| |  |
Db        755 TCAAAAACATGAAAGAACACATACTGGAGAGAAACCTTACAAGTGTGATCAATGTAATAA 814

Qy        360 AGCATTT---TATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
              |||  |||   ||| ||  | | |     |   ||| | | |  |  | |   || ||| |
Db        815 AGCCTTTGTATATGAAAGTTATTTACAAGTTCATAAAAAAACACATACTGGAGAGAAACC 874

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
               |  | |||||  |||||  |||||  ||  | ||||      |  |   |  |||||
Db        875 TTACAAATGTAATGAATGTGGTAAAGCCTTTGCACGACACAGTCATCTCAAAGTGCATAA 934

Qy        477 ATTAAAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
              | ||| ||||  ||| |      |   ||||  ||||| |||   ||||
Db        935 AATAACACATACTGGAGAGAAAC---CTTACAAATGTAATCAATGTGGTAAAGCCCTTGC 991

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
               |  || | |         | ||||       | | |         |         | |||||
Db        992 ATATCATAGTACACTCCAAGTACATCAAAGAACACATACTGGAGAGAAGCCCTATGAATG 1051

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
              |    ||||||| ||||  | ||       |||    |||      |||||    ||  ||
Db       1052 TGAGCAATGTGGCAAAGCCTTTGCAAATCAAAGTTATTTCCAAGTACATAAAAGAATACA 1111

Qy        657 TGATG 661
               |  ||
Db       1112 TACTG 1116


RESULT 6
```

```
BC047646
LOCUS       BC047646               2647 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone
            IMAGE:4828290), with apparent retained intron.
ACCESSION   BC047646
VERSION     BC047646.1  GI:29126804
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2647)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2647)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 90 Row: i Column: 7
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4828290"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN

  Query Match             7.9%;  Score 97.8;  DB 3;  Length 2647;
```

```
  Best Local Similarity   50.1%;  Pred. No. 4.4e-09;
  Matches  330;  Conservative    0;  Mismatches  317;  Indels   12;  Gaps    3;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | ||||         ||  ||     |  || |   |     |   | | |
Db       1235 GAATGTGGCAAAACTTTTAACCGATTCTCAACTCTTACTACCCATAAGATAATTCATACT 1294

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |    | || || | |||  |||| |||||| | ||| ||| | |||    |
Db       1295 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTT 1354

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                  |||  || ||  || | |    ||   ||| || ||||||      || || |||
Db       1355 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 1408

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||         ||| |  |    |||||||   ||   ||| | | | |
Db       1409 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 1468

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  |   ||||| ||||   ||||          ||||  | |  | |
Db       1469 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACA 1528

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||||  | |  |||||                |  | || |||||| |||  | |||| |||
Db       1529 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 1588

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||  |    ||  |   | |||| | ||| |   |  |   || ||| | |  |
Db       1589 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 1648

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              |||||| |||||  ||||| ||| |  |   ||||||    |  |  |||||||  |||
Db       1649 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 1708

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              |||  ||| |    |||  |      ||  |        ||||||  | | |  |
Db       1709 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 1768

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
              |     | | |    |   |  |  | | ||  |  | |||||  |||
Db       1769 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 1825

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||||| |||| || |      |      | | ||  ||||| |  |||||||   ||
Db       1826 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1884


RESULT 7
AU123448
LOCUS       AU123448                745 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU123448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
            sequence.
ACCESSION   AU123448
VERSION     AU123448.1  GI:10948164
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
            Suzuki,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RM2000319"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RM2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
ORIGIN

  Query Match             7.9%;  Score 97.6;  DB 1;  Length 745;
  Best Local Similarity   50.2%;  Pred. No. 4.4e-09;
  Matches  328;  Conservative    0;  Mismatches  314;  Indels   12;  Gaps    3;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | ||||         ||| ||     |  || |   |     |   | | |
Db         87 GAATGTGGCAAAACTTTTAACCGATCCTCAACTCTTACTACCCATAAGATAATTCATACT 146

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |    | || || | |||  |||| |||||| | ||| ||  | |||    |
Db        147 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGTTCTTCAACCCTT 206

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                  |||  || ||  || | |    ||   ||| || ||||||      || || |||
Db        207 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 260

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||         ||| |  |    |||||||   ||   ||| | | | |
Db        261 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 320

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  |   ||||| ||||   ||||          ||||  | |  | |
Db        321 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACTACA 380

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||||  | |  |||||                |  | || |||||| |||  | |||| |||
Db        381 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 440

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||  |    ||  |   | |||| | ||| |   |  |   || ||| | |  |
Db        441 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 500

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||||  |||||  ||||| ||| |  |   ||||||    |  |  |||||||  |||
Db        501 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 560

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              |||  ||| |      |||      ||   |        ||||||  | | |  |
Db        561 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 620

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
              |     | | |    |   |  |  | | ||  |  | |||||  |||
Db        621 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 677

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCAT 657
              ||||| |||| || |      |      | | ||  ||||| |  |||||||
Db        678 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCAT 731


RESULT 8
CO434402
LOCUS       CO434402                727 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  UI-M-HX0-csl-h-09-0-UI.r1 NIH_BMAP_HX0 Mus musculus cDNA clone
            IMAGE:30687440 5', mRNA sequence.
```

```
ACCESSION   CO434402
VERSION     CO434402.1  GI:49680696
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30687440"
                     /tissue_type="whole eye"
                     /dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16,
                     17, 18 dpc)"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_HX0"
                     /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AATAATTACG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH)."
ORIGIN
```

```
  Query Match             7.6%;  Score 93.4;  DB 7;  Length 727;
  Best Local Similarity   52.1%;  Pred. No. 3e-08;
  Matches  261;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          4 AGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGT 63
              | | |  |||| || ||   ||  ||    |    |   ||| |    | ||| |
Db         78 AATCAGTGTGATAAAGCCTTTTCTCGATACAGTACTCTTCAATCTCATAGAAGAACACAT 137

Qy         64 CCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTA 123
                |  || |    | || || | |   |||||||||||||| |  |    ||
Db        138 ACTGGAGAGAAACCCTTCAAATGTAATCAGTGTGATAAAGCCTTTTCTGAGAAATGTAGT 197

Qy        124 TTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGAT 183
                |  | |||| |||| |  || | |   ||   ||| || |||||||      |  |
Db        198 CTCCAAAACCATGTAAGGATACATACTGGAGAGAAACCCTACAAATGTA------ATCAA 251

Qy        184 TGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAA 243
              |||||||||| |||| |   || |  ||  |  | |  ||||    | ||||| ||| |
Db        252 TGTGATAAAGCCTTTTCCCATAACTATCACCTCCACATTCATAGAAGAACACATACCGGA 311

Qy        244 AAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAA 303
                | ||||| | | | |||     |||||  |||||   ||     |    | ||| |  |
Db        312 GAGAAACCCTACAAATGTAATCAGTGTGATAAAGCCTTTTCCCATAACTATCACCTCCAC 371

Qy        304 AGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCA 363
              |  |||  ||  || |||||              |  | || | |  | ||| |  ||||
Db        372 ATTCATAGAAGAACACATACTGGAGAGAAGCCTTACAAATGTAATCAGTGTGATAAAGCC 431

Qy        364 TTTTATAAACATCAATCTTTAAGACATCATATATTATC---TGTTCATGAAAAAACATTA 420
              |||| | || |    ||| |   |  ||| | |  | |   |  |   || ||| ||||
Db        432 TTTTCTCAATACTTTTCTCTTCAAACTCACAGAAGAACACATACTGGAGAGAAACCATTC 491

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              |  ||||| |||||| ||||||| ||  ||| |       |   |   |  |||| | |
Db        492 AAATGTAATCAATGTGATAAAGTCTTTTCTCAATACAGTACTCTTCAAACTCATAGAAGA 551

Qy        481 AAACATCATGGTGGATCTCCT 501
              | ||||  ||| |     |||
Db        552 ACACATACTGGAGAGAAACCT 572
```

```
RESULT 9
CN352904
LOCUS       CN352904                631 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000600057466 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN352904
VERSION     CN352904.1  GI:47352838
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 631   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, DMSO-treated H9 cell
                     line"
                     /clone_lib="GRN_PREHEP"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from DMSO-treated hES cell line H9 (p22) maintained in
                     feeder-free conditions"
ORIGIN
```

```
  Query Match             7.5%;  Score 92.4;  DB 7;  Length 631;
  Best Local Similarity   54.1%;  Pred. No. 4.7e-08;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   || || || | |||  ||||| ||||| | ||| | | | |||  | |
Db         83 AGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAAACTTAC 142

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                |||||    | ||  || | |   ||   ||| ||| |||||       ||  | ||||
Db        143 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTATGAATGT------GAAAAATGTGG 196

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  |       |||  || |    | |||||     |   ||| | ||| | ||
Db        197 CAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGAAGAGAA 256
```

```
Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  |||||  |||||  ||||  |  |     |  | | |  |  ||
Db        257 ACCTTACAAATGTGAAGAATGTGGCAAAGGTTTTAAATGGCCCTCAACCCTTACTATCCA 316

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | ||  |  |||||              ||  | ||  ||||  | |||     ||||  ||| |
Db        317 TAAGATAATTCATACTGGAGAGAAACCATACAAATGTGAAGAATGTGGCAAAGCTTTTAA 376

Qy        369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                ||       ||  |   | |||| | | | |    |  |   || ||| | |  || ||
Db        377 CCAATCCTCAAAACTTACCAAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATG 436

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   ||||  || |  |   | |||||  |  |  |||||||   ||  ||
Db        437 TGAAGAATGTGGCAAAGCCTTTAACCAGTCCTCAAACCTTACTAAACATAAGAGAATTCA 496

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| | |   |||
Db        497 TACTGGAGAAAAACCT 512
```

RESULT 10
BC036394
LOCUS BC036394 2672 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone IMAGE:5259399), containing frame-shift errors.
ACCESSION BC036394
VERSION BC036394.1 GI:23025784
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2672)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2672)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 73 Row: a Column: 16
This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
source 1. .2672
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5259399"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
ORIGIN

Query Match 7.5%; Score 92.4; DB 3; Length 2672;
Best Local Similarity 54.1%; Pred. No. 5.2e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   || || || | |||  ||||| ||||| | ||| | | | |||  | |
Db       1745 AGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAAACTTAC 1804

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                |||||   | ||  || | |   ||   ||| ||| |||||      ||  | ||||
Db       1805 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTATGAATGT------GAAAAATGTGG 1858

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               ||||| |||  |    ||| || |   | |||||   |  ||| | ||| | ||
Db       1859 CAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGAAGAGAA 1918

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  |||||  |||||  ||||  |  |     |  | | |  |  ||
Db       1919 ACCTTACAAATGTGAAGAATGTGGCAAAGGTTTTAAATGGCCCTCAACCCTTACTATCCA 1978

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | ||  |  |||||              ||  | ||  ||||  | |||     ||||  ||| |
Db       1979 TAAGATAATTCATACTGGAGAGAAACCATACAAATGTGAAGAATGTGGCAAAGCTTTTAA 2038

Qy        369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                ||       ||  |   | |||| | | | |    |  |   || ||| | |  || ||
Db       2039 CCAATCCTCAAAACTTACCAAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATG 2098

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   ||||  || |  |   | |||||  |  |  |||||||   ||  ||
Db       2099 TGAAGAATGTGGCAAAGCCTTTAACCAGTCCTCAAACCTTACTAAACATAAGAGAATTCA 2158

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| | |   |||
Db       2159 TACTGGAGAAAAACCT 2174
```

RESULT 11
AQ005136
LOCUS AQ005136 591 bp DNA linear GSS 27-JUN-1998
DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017, genomic survey sequence.
ACCESSION AQ005136
VERSION AQ005136.1 GI:3082581
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 591)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

```
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2290O17.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7150844"
                     /db_xref="taxon:9606"
                     /clone="2290O17"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.4%;  Score 92;  DB 8;  Length 591;
  Best Local Similarity   52.2%;  Pred. No. 5.6e-08;
  Matches  256;  Conservative    0;  Mismatches  225;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |     || ||     |  || |   |     |   | | |
Db         69 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 128

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | || |    | || || |||||  |||| |||||| | ||| |  || |||    |
Db        129 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 188

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                ||||       | ||   ||| ||   ||| || ||||||      || || |||
Db        189 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTCACAAATGT------GAAGAATGT 242

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |    | |||||   ||   ||||| | | |
Db        243 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 302

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||  |   ||||| |||||  |||| |       |  | | | || |
Db        303 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 362

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| |||  |  |||||               | | || |||| | |||    |||| |||
Db        363 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 422

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               ||     |   ||   |  | |||| | ||  |   |  |   || ||| | | |
Db        423 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 482

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |  |   | |||||| |  |   ||||||  |||
Db        483 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 542

Qy        484 CATCATGGTG 493
              |||  ||| |
Db        543 CATACTGGAG 552


RESULT 12
CN362485
```

```
LOCUS       CN362485                687 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000424188118 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN362485
VERSION     CN362485.1  GI:47362419
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 687   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, embryoid bodies
                     derived from H1, H7 and H9 cells"
                     /clone_lib="GRN_EB"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from embryoid body outgrowths derived from hES cell lines
                     H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                     conditions."
ORIGIN

  Query Match             7.4%;  Score 92;  DB 7;  Length 687;
  Best Local Similarity   52.2%;  Pred. No. 5.6e-08;
  Matches  256;  Conservative    0;  Mismatches  225;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||      || ||      |     |   |  |
Db        183 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 242

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               |  |  | || | | || || |||||  | ||  | |||||  |||  |  | |||     |
Db        243 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 302

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |  |||||   ||||  || | |   ||   ||| || ||||||| |      || |||
Db        303 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 356

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |   |  |||||  |     | |||||     |   ||| | ||| |
Db        357 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 416

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  ||||| ||||    ||||      |   | | |  | |
Db        417 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 476

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||    |  | ||    |  | || ||| || |||      ||| |||
Db        477 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 536

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
               |  ||      |    | ||||  |||  |  |       || |||   |
Db        537 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 596

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||    |||| ||| |  ||| | |||||| |  |  |||||||  | |
Db        597 TGTGAAGAATGCGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGTACATT 656
```

```
Qy        484 CATCATGGTG 493
              ||| ||| |
Db        657 CATACTGGAG 666


RESULT 13
CN417582
LOCUS       CN417582               752 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000424010604 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN417582
VERSION     CN417582.1  GI:47405176
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 752   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, embryoid bodies
                     derived from H1, H7 and H9 cells"
                     /clone_lib="GRN_EB"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from embryoid body outgrowths derived from hES cell lines
                     H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                     conditions."
ORIGIN

  Query Match             7.4%;  Score 92;  DB 7;  Length 752;
  Best Local Similarity   52.0%;  Pred. No. 5.7e-08;
  Matches  259;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |     || ||     |  || |   |     |   | | |
Db         72 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 131

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
             | || |    | || || |||||  |||| ||||||  | ||| |  || |||    |
Db        132 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 191

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |    ||||      |  ||   ||| ||   ||| || ||||||      || || |||
Db        192 GCTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 245

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |     | |||||   ||   |||| | | |
Db        246 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 305

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | |||   |   ||||| |||||  |||| |         | | | | || |
Db        306 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 365

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  | |||||              | | || |||| | |||    |||| |||
Db        366 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 425

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||     |   ||   |  | |||| | ||  |   |  |   || ||| | | |
Db        426 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 485

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| || | |  | |||||| |  |   |||||| |||
Db        486 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 545

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |    |||
Db        546 CATACTGGAGAGAAACCT 563


RESULT 14
B99387
LOCUS       B99387                 593 bp    DNA     linear   GSS 26-JUN-1998
DEFINITION  CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
            genomic survey sequence.
ACCESSION   B99387
VERSION     B99387.1  GI:3027197
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7147044"
                     /db_xref="taxon:9606"
                     /clone="2281A10"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.4%;  Score 91;  DB 8;  Length 593;
  Best Local Similarity   54.4%;  Pred. No. 8.8e-08;
  Matches  231;  Conservative    0;  Mismatches  185;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              |||| |    | || || |||||  ||||  ||||| | ||  |    | |||     |
Db         47 AAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCAGTCCTCAACCCTTAG 106

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
              |||||   | ||  || | |   ||   ||| || ||||||      || || ||||
Db        107 AAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 160

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||| |||    |   |||||  |    | |||||  |||   ||| | ||| | ||
```

```
Db        161 TAAAGCTTTTAAGTGGTCCTCACACCTTACTAGACATAAAGTAATTCATACTGAAGAGAA 220

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  ||||| || |  ||||  |     |  | | |  | |||
Db        221 ACCCTACAAATGTGAAGAATGTGGCAAGGCTTTTAACCATTTCTCAGCCCTTAGGAAACA 280

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || |  |||||           |  | || |||| | |||    |||| |||
Db        281 TAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAG 340

Qy        369 TAAACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                |      | ||  |   || |||| | ||| |   |  |   || ||| | |  |  ||
Db        341 CCAGTCCTCAACTCTTAGAAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATG 400

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||  |||||  ||| |   |  | ||||||  |  |   ||||||  |||  ||
Db        401 TGAAGAATGTGGTAAAGCTTTTAAGTGGTCCTCAAAACTTACTGTACATAAGGTAATTCA 460

Qy        486 TCATG 490
              |  ||
Db        461 TACTG 465


RESULT 15
BX414627
LOCUS       BX414627                820 bp    mRNA    linear   EST 01-MAY-2004
DEFINITION  BX414627 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YD12
            5-PRIME, mRNA sequence.
ACCESSION   BX414627
VERSION     BX414627.2  GI:46930382
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 13, 2003 this sequence version replaced gi:30649903.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 9703.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0CAP001DB06QP1&c=9703.f.
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YD12"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             7.4%;  Score 91;  DB 5;  Length 820;
  Best Local Similarity   50.1%;  Pred. No. 9e-08;
  Matches  289;  Conservative    8;  Mismatches  268;  Indels   12;  Gaps    3;

Qy         89 AAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACA 148
              |||  |||| |||||| | ||| ||| | |||    |       |||  |:|::  || |

Db          2 AAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTACTACSCATAGARGRRTTCATA 61

Qy        149 GTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAAT 208
              |   ||   ||| || ||||||       || || ||:|  ||||| |||          |
Db         62 CTGGAGAGAAACCTTACAAATGT------GAAGAATGKGGCAAAGCCTTTAAGCAGTCCT 115

Qy        209 CACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGT 268
              || |  |    |||||||   ||   ||| | | | | ||||| | | | |||  |   |
Db        116 CAAACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAAGAT 175

Qy        269 GTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGT 328
              |||| ||||   ||||          |||| | |  | ||||||  | |  |||||
Db        176 GTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACATGAGGTAATTCATACTGGAG 235

Qy        329 CATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACATCAATCTTTAA 385
                    |  | || |||||  |||  | |||| ||| || |       ||   |   |
Db        236 AGAAACCCTACAAATGTGAAAGATGTGGAAAAGCCTTTAATCATTTCTCACACCTTACTA 295

Qy        386 GACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTT 445
              |||| | ||| |   |  |   || ||| | |   | ||||||  :|||  ||||| ||
Db        296 CACATAAGATAATTCATACTGGAGAGAAACCTTACAAATGTAAAGRATGTGGTAAAGCTT 355

Qy        446 TCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTT 505
              | |  |   ||||||    |  |  |||||||  |||  |||  |||  |      |||
Db        356 TTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCATACTGGAGAGAAGCCTTACA 415

Qy        506 ATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTC 565
              |       |:   |    |  ||||||  | | |  ||       | | |    |
Db        416 AATGTAAAGRATGTGAAAGAGCTTTTAACCAATCCTCAAAACTTACTGAACATAAGARAA 475

Qy        566 ATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGA 625
              |   |     | | ||  |   | |||||   |:|||||  |||| || |     |
Db        476 TTCATACTGGAGAGAAACCCTA---TGAATGTGAAARATGTGGCAAAGCTTTTAACCAGT 532

Qy        626 AAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
                |  | |  ||  ||||| |  :||||||   ||
Db        533 CCTCAAATCTTACTAGACATAAGAGRAGTCATACAGA 569


Search completed: February  9, 2005, 07:43:36
Job time : 3816.38 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:51:59 ; Search time 5292.91 Seconds
(without alignments)
11342.728 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1239	100.0	1239	6	BD274350	BD274350 Candida a
2	1239	100.0	2060	6	BD274349	BD274349 Candida a
3	1235.8	99.7	1239	6	AX489193	AX489193 Sequence
4	584.8	47.2	627	6	AR550053	AR550053 Sequence
c 5	438.4	35.4	110000	8	CR382135_06	Continuation (7 of
6	391	31.6	462	6	AR550055	AR550055 Sequence
c 7	228	18.4	110000	8	CR382122_03	Continuation (4 of
c 8	227.2	18.3	92211	8	CR380953_09	Continuation (10 o
9	217.4	17.5	1560	8	YSCNOFEAT	M90638 Saccharomyc
10	217.4	17.5	1739	8	YSCTFIIIA	M80611 Saccharomyc
c 11	217.4	17.5	37497	8	YSCP9677	U25841 Saccharomyc
c 12	199.2	16.1	2133	8	YSCRP026A	M33924 S.cerevisia
13	193	15.6	797	11	CNS06ING	AL400514 T7 end of
14	120.6	9.7	110000	8	AE016818_01	Continuation (2 of
c 15	101	8.2	169063	9	AC138126	AC138126 Homo sapi
c 16	101	8.2	177299	9	AC073544	AC073544 Homo sapi
17	99.4	8.0	2132	6	AX714042	AX714042 Sequence
18	99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
19	99.4	8.0	2196	6	CQ726027	CQ726027 Sequence
c 20	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
c 21	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
22	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
23	97.8	7.9	1203	9	D70831	D70831 Homo sapien
c 24	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
c 25	97.2	7.8	115995	9	AC011494	AC011494 Homo sapi
26	96.6	7.8	2320	6	AR270491	AR270491 Sequence
27	96.6	7.8	2320	9	HSU35376	U35376 Human repre
c 28	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
c 29	96.6	7.8	106811	2	CR556705	CR556705 Mus muscu
30	95	7.7	981	9	HSHZF9	X78932 H.sapiens H
c 31	94.6	7.6	186233	9	AC092329	AC092329 Homo sapi
32	94.6	7.6	189317	2	AC024483	AC024483 Homo sapi
33	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
34	93.6	7.6	1389	9	M27879	M27879 Homo sapien
c 35	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
36	93.4	7.5	199585	10	AC124426	AC124426 Mus muscu
37	92.4	7.5	3516	6	CQ726030	CQ726030 Sequence
c 38	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
39	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
40	92	7.4	2873	9	HUMHPLK	M55422 Human Kruep
41	92	7.4	2873	11	G28705	G28705 sWSS3976 Er
42	92	7.4	2873	11	HUMSWS1269	G18281 human chrom
43	92	7.4	156835	9	AC010620	AC010620 Homo sapi
c 44	92	7.4	191082	2	AC141066	AC141066 Homo sapi
c 45	92	7.4	203396	9	AC073210	AC073210 Homo sapi

ALIGNMENTS

RESULT 1
BD274350
LOCUS BD274350 1239 bp DNA linear PAT 17-JUL-2003
DEFINITION Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA protein.
ACCESSION BD274350
VERSION BD274350.1 GI:33084118
KEYWORDS JP 2002531068-A/2.
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1239)
AUTHORS Pallier,F.B., Camier,S. and Sentenac,A.
TITLE Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
JOURNAL Patent: JP 2002531068-A 2 24-SEP-2002;
AVENTIS PHARMA SA
COMMENT OS Candida albicans
PN JP 2002531068-A/2
PD 24-SEP-2002
PF 09-NOV-1999 JP 2000581204
PR 10-NOV-1998 FR 98/14147
PI FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC C07K16/14,
PC C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC G01N33/53,
PC G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
PC 00,
PC (C12N15/00,C12R1:725)
CC Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
CC protein
FH Key Location/Qualifiers
FT CDS (1). .(1236).
FEATURES Location/Qualifiers
source 1. .1239
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"
ORIGIN

Query Match 100.0%; Score 1239; DB 6; Length 1239;

```
  Best Local Similarity   100.0%;  Pred. No. 5.7e-190;
  Matches 1239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              ||||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239


RESULT 2
BD274349
LOCUS       BD274349               2060 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            protein.
ACCESSION   BD274349
VERSION     BD274349.1  GI:33084117
KEYWORDS    JP 2002531068-A/1.
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 2060)
  AUTHORS   Pallier,F.B., Camier,S. and Sentenac,A.
  TITLE     Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
  JOURNAL   Patent: JP 2002531068-A 1 24-SEP-2002;
            AVENTIS PHARMA SA
COMMENT     OS   Candida albicans
            PN   JP 2002531068-A/1
            PD   24-SEP-2002
            PF   09-NOV-1999 JP 2000581204
            PR   10-NOV-1998 FR     98/14147
            PI   FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC
            C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC
            C07K16/14,
            PC   C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC
            G01N33/53,
            PC   G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
            PC   00,
            PC   (C12N15/00,C12R1:725)
            CC   Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            CC           protein
            FH   Key             Location/Qualifiers
            FT   source          1. .2060
            FT                   /organism='Candida albicans'.
FEATURES             Location/Qualifiers
     source          1. .2060
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match           100.0%;  Score 1239;  DB 6;  Length 2060;
  Best Local Similarity   100.0%;  Pred. No. 5e-190;
  Matches 1239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899
```

```
Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1958
```

```
RESULT 3
AX489193
LOCUS       AX489193               1239 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6493 from Patent WO02053728.
ACCESSION   AX489193
VERSION     AX489193.1  GI:22323205
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6493 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1239
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match             99.7%;  Score 1235.8;  DB 6;  Length 1239;
  Best Local Similarity   99.8%;  Pred. No. 1.9e-189;
  Matches 1237;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239


RESULT 4
AR550053
LOCUS       AR550053                627 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5184 from patent US 6747137.
ACCESSION   AR550053
VERSION     AR550053.1  GI:53943228
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5184 08-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             47.2%;  Score 584.8;  DB 6;  Length 627;
  Best Local Similarity   97.4%;  Pred. No. 1.3e-84;
  Matches  608;  Conservative    0;  Mismatches    7;  Indels    9;  Gaps    1;

Qy        616 TGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATC 675
              ||| | || ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         13 TGTTTGGGAAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTTACCATGATC 72

Qy        676 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 132

Qy        736 GAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAA 795
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        133 GAACATTATAATATCTTCCATGATGGTAATATCCCCGATGATTTATTAAAGGAAACTGAA 192

Qy        796 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 252

Qy        856 GAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAAGATAGTCTAGATGAA 915
              |||||||||||||||||||||         ||||||||||||||||||||||||||||||
Db        253 GAAACAGAGAAATTAAAAGTG---------GAAGAAGATGAAGAAGATAGTCTAGATGAA 303

Qy        916 AAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 975
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        304 AGAAGAAGTGATGTTAGATCAGATTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 363

Qy        976 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 423

Qy       1036 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 483

Qy       1096 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 543

Qy       1156 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 603

Qy       1216 GAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||||
Db        604 GAAACATCAGTGATTTCTCGATAA 627


RESULT 5
CR382135_06/c
WPCOMMENT
  Sequence split into 16 fragments  LOCUS CR382135 Accession CR382135
     Fragment Name      Begin     End
     CR382135_00            1    110000
     CR382135_01       100001    210000
     CR382135_02       200001    310000
     CR382135_03       300001    410000
     CR382135_04       400001    510000
     CR382135_05       500001    610000
     CR382135_06       600001    710000
     CR382135_07       700001    810000
     CR382135_08       800001    910000
     CR382135_09       900001   1010000
     CR382135_10      1000001   1110000
     CR382135_11      1100001   1210000
     CR382135_12      1200001   1310000
     CR382135_13      1300001   1410000
     CR382135_14      1400001   1510000
     CR382135_15      1500001   1592360
Continuation (7 of 16) of CR382135 from base 600001 (CR382135 Debaryomyces hansenii chro

  Query Match             35.4%;  Score 438.4;  DB 8;  Length 110000;
  Best Local Similarity   63.4%;  Pred. No. 1.4e-61;
  Matches  724;  Conservative    0;  Mismatches  406;  Indels   12;  Gaps    3;

Qy         33 ATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACATATGAAGG 92
              ||| | ||   | || || |||||| | || || ||||| ||||||||     |   ||
Db      20471 ATCCTCAAGGGCGTCATCGTCTTCAGCTCGACCTAAAAATTATATTTGTGACGAGCCAGA 20412

Qy         93 GTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACAGTAA 152
              || |  || |||||  |  ||||||  | ||||| ||||||  |||| | || | |||
Db      20411 CTGCGGGAAGGCCTATTCTAAACCATCGCTCTTAGAACAACATAAAAGATCACATACTAA 20352

Qy        153 TGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACA 212
              |||  |||| | ||||||  | |     |  || || ||  ||||  | ||||| |||||
```

```
Db      20351 TGAAAGACCCTTTAAATGCTCGGAAACTGGATGCGACAAGTCATTCCTAAGAAAGTCACA 20292

Qy        213 TTTGGAAACACATATTGTATC------ACATTCCGAAAAAAAACCATTCCATTGTTCAGT 266

Db      20291 TTTACAAGCACATTTGCTTTCTCATGAAGATCAAGAAAGTAAACCCTTCCAGTGTGCAAC 20232

Qy        267 GTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAA 326

Db      20231 TTGCGGCAAAGGTGTAAATACCTTACAACATTTAAAGAGACACGAAATAACACATACAAA 20172

Qy        327 GTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAG 386

Db      20171 ATCATTCATTTGTACATTTGAGGGCTGTAATGAATCATTTTATAAGCATCAGTCACTCAG 20112

Qy        387 ACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTT 446

Db      20111 ACACCATACGCTATCGGTTCATGAAAAAAAATTATTATGCAATAAGTGTAATAAATCGTT 20052

Qy        447 CACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTTA 506

Db      20051 TAACAGACCGTATCGATTGGCACAGCACAATATTAAATATCACAGTGACTCGCCGGCTTA 19992

Qy        507 TCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCA 566

Db      19991 TCAATGTGACCATCAAGGCTGTTTTGGAAATTTTATGACGTGGTCGGCTTTACAATTGCA 19932

Qy        567 TATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAA 626

Db      19931 TATAAAAACTGAGCACCCAAAGATAAAGTGTCCTATATGTGGGAAAGGTTGCGTGGGTAA 19872

Qy        627 AAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGAC 686

Db      19871 GAAAGGGCTTCGGTCACACATGAATATTCACGATGAAGAAAAGATAGTCAAATTATGGAA 19812

Qy        687 TTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTATAA 746

Db      19811 TTGTAATTATTGTAACATTGGAAAATTTGTGAAAAAAGCTGATTTGATAGATCATTACAA 19752

Qy        747 TATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAAGTGAAAAAATT 806

Db      19751 CACTTATCACGACAGAAATCTTCCTGATGATTTATTGAAACCGATCGAACGAGAACAGTT 19692

Qy        807 AGAGAACCTATTA---GATCAAGGATCGAAATTAAATAATTTGCATGAATTAGAAACAGA 863

Db      19691 AGATAAACTTCTAAGCGAGAAGGACTCTACCAATATTGATTTTAATACTTTAGAAGGATT 19632

Qy        864 GAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAAGATAGTCTAG---ATGAAAAAAG 920

Db      19631 ACAATCAAAAGGGTTTGTGGAAGTACCGTCGGATGAAGAGAAGGAAGAGAATGATTACAT 19572

Qy        921 AAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCAATAAAATCATTTACTGCTTC 980

Db      19571 GTCCAATGGTTTGCAAGCTTCACATAAATCTTTGAATTCGTTGAATTTAACTTTGGAATC 19512

Qy        981 TTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATCAATTG 1040

Db      19511 TGGAAAAGCATCTATTGTAGATCTTATTCTGAATAATTATCTGAAAAGGAAAGTTCCATG 19452

Qy       1041 TCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACATTTGAA 1100

Db      19451 TCCGAAAAAGAATTGTGATAGAATGTTCAGCCGTGATTATGATCTTGAGAGACATTTAAA 19392

Qy       1101 ATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAAGAAGA 1160

Db      19391 GTGGCATGAATCTCACTTAAAAAAGATTGAAGCGTTTTTAAAAAGTCTTGAGCTGGAAGA 19332

Qy       1161 AA 1162

Db      19331 AA 19330
```

RESULT 6
AR550055

```
LOCUS       AR550055                462 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5186 from patent US 6747137.
ACCESSION   AR550055
VERSION     AR550055.1  GI:53943230
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5186 08-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           31.6%;  Score 391;  DB 6;  Length 462;
  Best Local Similarity   98.5%;  Pred. No. 2.1e-53;
  Matches  405;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Db         52 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 111

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Db        112 CGTCCCAAAAAGTATATTTGTACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 171

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180

Db        172 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 231

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Db        232 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 291

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Db        292 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 351

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Db        352 AAAAGACATGAAATCACCCATACCAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 411

Qy        361 GCATTTTAT-AAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410

Db        412 ACATTTTATAAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 462
```

```
RESULT 7
CR382122_03/c
WPCOMMENT
  Sequence split into 14 fragments  LOCUS CR382122 Accession CR382122
     Fragment Name      Begin     End
     CR382122_00            1    110000
     CR382122_01       100001    210000
     CR382122_02       200001    310000
     CR382122_03       300001    410000
     CR382122_04       400001    510000
     CR382122_05       500001    610000
     CR382122_06       600001    710000
     CR382122_07       700001    810000
     CR382122_08       800001    910000
     CR382122_09       900001   1010000
     CR382122_10      1000001   1110000
     CR382122_11      1100001   1210000
     CR382122_12      1200001   1310000
     CR382122_13      1300001   1320834
Continuation (4 of 14) of CR382122 from base 300001 (CR382122 Kluyveromyces lactis stra
```

```
  Query Match             18.4%;  Score 228;  DB 8;  Length 110000;
  Best Local Similarity   58.1%;  Pred. No. 9.8e-28;
  Matches  443;  Conservative    0;  Mismatches  310;  Indels    9;  Gaps    2;

Qy         41 TATCTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATA 100
              | ||   ||| || |||   ||| | |||   ||| | |||   ||| ||||||| | |
Db      15150 TGTCAGTTTCATCGTCACAGCGTTCTAAAGTATATTTCTGCGATTATCAAGGGTGTAACA 15091

Qy        101 AAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGAC 160
              || | |   ||||    ||  |   | | |||       ||   |    || || ||
Db      15090 AATCGTTCACAAGACCGAGTTTGCTCACGGAGCATCAGCTCAC---CGTGCATCATGGAA 15034

Qy        161 CGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAA 220
              |   | || |   |||  |||||  |||| ||||   | |||||| ||| || | |
Db      15033 TCAAACCGTTTAAATGCGACACTTGTGGCAAAGAATTTGCGAAAAAATCGCATCTGAATA 14974

Qy        221 CACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGG 280
              ||| ||  | || ||| | ||  | ||||| |||   |||||  | ||||| ||||| |
Db      14973 GACACATGTTCTCTCATACAGATGATAAACCTTTCACATGTTCTATATGTGGGAAAGGTG 14914

Qy        281 TTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTA 340
              | |    | ||||||| || || |||||||||| || ||||| ||||| ||  | |||
Db      14913 TCACAACCAGACAACAGTTAAAGAGACATGAAATTACACATACGAAGTCCTTCCACTGTT 14854

Qy        341 CATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTAT 400
              ||| |||    ||| | ||| |||| || ||||||| |   |   ||  |||||  || |
Db      14853 CATATGAGGGATGTAACGAATCATTCTACAAACATCCACAACTCCGATCGCATATCTTGT 14794

Qy        401 CTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAA 460
              | ||||| ||||| | | | || ||    || ||||| ||    |||    | ||||
Db      14793 CCGTTCACGAAAAGAAACTTACCTGCCCTCACTGTAACAAGACGTTCCAAAGGCCTTATC 14734

Qy        461 AATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGC------TTATCAATGTG 514
              |||      | || | |   || |||||       |    || |       || |||||||
Db      14733 GATTGAAGAATCACATAGATAAGCATCACAACCCAGAATCTACCGGAATGTACCAATGTG 14674

Qy        515 ATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAAC 574
              |   || |  |||    |   ||   |||||||    ||||||  ||||| ||||
Db      14673 ACTTTCTTTCATGTACCGACGTATTTTCCACTTGGTCTTCTTTACAACAACATATTAAAC 14614

Qy        575 AACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTT 634
              ||   ||||| ||| |   |||||||  |||||| |||  ||||||||    ||  |||||
Db      14613 AATGCCATCCGAAATTACCATGTCCTATTTGTGGGAAACCTTGTGTTGCTGAATCAGGTT 14554

Qy        635 TATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATT 694
              ||      |||||| | | |||||| ||    |   | | || |  ||||  |||  ||
Db      14553 TACGGAATCATATGATGATTCATGACGAGAGTTTGGTCACTAAGAATTGGAAATGTTCTT 14494

Qy        695 ATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCC 754
                |||   |       |||||| ||||||| | |  ||||  ||||||||  |
Db      14493 CATGTCCAGATACATCATTTGCGAAGAAAACTCAGCTAGTACAACATTATGAAGAAAGTC 14434

Qy        755 ATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAAG 796
              |  | | |       || ||      |||||  |    || || |
Db      14433 ACAAAGATGTAGATCCAGAATTGCTATTACTGCCGACGGACG 14392


RESULT 8
CR380953_09/c
WPCOMMENT
  Sequence split into 10 fragments  LOCUS CR380953 Accession CR380953
     Fragment Name      Begin     End
     CR380953_00            1    110000
     CR380953_01       100001    210000
     CR380953_02       200001    310000
     CR380953_03       300001    410000
     CR380953_04       400001    510000
     CR380953_05       500001    610000
     CR380953_06       600001    710000
     CR380953_07       700001    810000
     CR380953_08       800001    910000
     CR380953_09       900001    992211
Continuation (10 of 10) of CR380953 from base 900001 (CR380953 Candida glabrata strain C
```

```
  Query Match             18.3%;  Score 227.2;  DB 8;  Length 92211;
  Best Local Similarity   57.8%;  Pred. No. 1.4e-27;
  Matches  447;  Conservative    0;  Mismatches  318;  Indels    9;  Gaps    2;

Qy         44 CTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAG 103
              | |||   ||      | |  | ||| |  ||| ||||   ||| ||    ||||| || |
Db      60262 CATCTGGATCCAGTGTAAGGTCGAAGACATATCTTTGTGAATACGATAATTGTGACAAGG 60203

Qy        104 CCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGT 163
              ||| ||   | || || ||  |     ||||    |  ||         ||  | |   |
Db      60202 CCTTTACAAGGCCGTCCTTGCTCACAGAACACCAGAATAC---AGTGCATCTTGGGAGGA 60146

Qy        164 ATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACAC 223
              |    || | |    | | ||||| |   |||||    | |||     || || || | ||
Db      60145 AACCTTGGAAATGTAATCAATGTGAAAGTTCATTTACTAAAAAGATCCACTTAGAGAGAC 60086

Qy        224 ATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTA 283
              |  |      ||||  | ||  ||| ||| ||  |||||||| | ||||| || ||| | |
Db      60085 ACTTATACACACACACAGATGAAAGACCGTTTTATTGTTCATTTTGTGGGAAGGGGCTAA 60026

Qy        284 ATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACAT 343
              | || | |||||  ||||| |||||||  | || || || || || || || ||| ||
Db      60025 TTACTAGGCAACAACTGAAACGACATGAGGTAACTCACACCAAATCGTTCAATTGTGAAT 59966

Qy        344 TTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTG 403
              ||||  |||| | ||     || || || || |      |||||   |||||| ||| | |
Db      59965 ATGAAGGTTGTAATGAGAGTTTCTACAAGCACCCTCAATTAAGGGCTCATATTTTAGCAG 59906

Qy        404 TTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAAT 463
              |||||  | |||   ||| ||| |  |||| || |||  |||      |||||  | |
Db      59905 TTCATTTACAAAGTCTAAAATGTCACGAATGCAACAAAAGTTTTCAAAGACCTTACAGAC 59846

Qy        464 TAGCACAACATAAATTAAAACATCAT------GGTGGATCTCCTGCTTATCAATGTGATC 517
              |   | || || | |   ||||| |||      | ||          ||||||||||||
Db      59845 TCAAAAATCACATAGCCAAACACCATAATCCAGATGTTGTCAATGCTTATCAATGTACCT 59786

Qy        518 ATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAAC 577
              || ||||| ||| | |||| |||| ||||| |  |||  ||| ||| | ||| |
Db      59785 TCAGTGTTTGTTCTAAGAGTTTCAAAACATGGTCTGCGTTAAGATTACATGTTAAAAATG 59726

Qy        578 TGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTAT 637
              |||||  ||  | ||||| |||| |||  |||||  |||||||| || | ||||||
Db      59725 ATCATCCGAAGTTAAAATGCCCTATATGCAGTAAACCCTGTGTTGGGGAAGACGGTTTAA 59666

Qy        638 CTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATT 697
              |   || |||  ||  || |||||        |    | ||  ||||  ||| ||   |
Db      59665 ATATGCACATGAAAATCCACGATGAAAACCTTGTTTCAAGAAATTGGAAGTGTCATATAT 59606

Qy        698 GTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCCATG 757
              || |||     | |||||| ||||||   ||| || | || ||||| |  | |   |||
Db      59605 GTAATGATCAGTCATTTGCTAAGAAATTAGAACTATTGGATCATTACAGCAACAGTCATA 59546

Qy        758 ATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAAGTGAAAAAATTAGAGA 811
              ||   | ||||| |   || ||||| |  | |  | ||||     |  | || |
Db      59545 GTGAAGAAATCCCGGCCTATCTATTAGAACAGAAGGTAGTGCCTTATGTTGAAA 59492


RESULT 9
YSCNOFEAT
LOCUS       YSCNOFEAT               1560 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  Saccharomyces cerevisiae transcription factor IIIA and RNA
            polymerase subunit RPB6 genes.
ACCESSION   M90638
VERSION     M90638.1  GI:172046
KEYWORDS    RNA polymerase; transcription factor IIIA.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Woychik,N.A. and Young,R.A.
  TITLE     Genes encoding transcription factor IIIA and the RNA polymerase
```

```
            common subunit RPB6 are divergently transcribed in Saccharomyces
            cerevisiae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)
  MEDLINE   92237295
   PUBMED   1570325
COMMENT     Original source text: Saccharomyces cerevisiae (strain S288C) DNA.
FEATURES             Location/Qualifiers
     source          1. .1560
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C"
                     /db_xref="taxon:4932"
ORIGIN      chromosome 14.

  Query Match             17.5%;  Score 217.4;  DB 8;  Length 1560;
  Best Local Similarity   59.4%;  Pred. No. 1.4e-25;
  Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
              | || || |  ||| | ||    ||||| || |||||||| || | ||   |||| |||
Db        368 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 427

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
              | ||     |||   |  || |     | | | |  |         || |      ||
Db        428 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 484

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
              | ||||  ||| | ||  | | |||    || || || | |||  |     | ||||| |
Db        485 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 544

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
              | |  |||||||||| ||||    ||||| ||||| || |   |||| || ||  |||
Db        545 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 604

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
              |  |||| ||| | || ||||| || || || |  ||| ||   |||   ||  |
Db        605 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 664

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
               ||| || || |||| |   |||||    ||||| ||||||||||||  | | | | |||
Db        665 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 724

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
              | |||  ||| || ||||||    ||      | || |  |   ||  | | || |  |
Db        725 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 784

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
              |||||||      ||  | |    | |||| ||||||  |  | |||||||||  |||
Db        785 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 844

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
              | || |  |  ||||||  ||| ||||  |||||||| |    |||||  ||| | ||||
Db        845 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 904

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
              |||||| ||| | |||   ||||| ||| |||| ||||||      || ||| | | ||
Db        905 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 964

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
              |||| || ||    | | ||||| ||||  ||| ||  |||  ||   |   |||
Db        965 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1024

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
              | |  ||| ||||  |  |    ||||||
Db       1025 CTAGAAAACATGATCTTCTCACGCATTAT 1053


RESULT 10
YSCTFIIIA
LOCUS       YSCTFIIIA               1739 bp    DNA     linear   PLN 13-SEP-1996
DEFINITION  Saccharomyces cerevisiae transcription factor IIIA (TFIIIA) gene,
            complete cds.
ACCESSION   M80611
VERSION     M80611.1  GI:172902
KEYWORDS    transcription factor IIIA; zinc-finger protein; zinc-finger
            transcription factor.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1739)
  AUTHORS   Archambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D.
            and Segall,J.
  TITLE     The deduced sequence of the transcription factor TFIIIA from
            Saccharomyces cerevisiae reveals extensive divergence from Xenopus
            TFIIIA
  JOURNAL   J. Biol. Chem. 267 (5), 3282-3288 (1992)
  MEDLINE   92147684
   PUBMED   1737784
FEATURES             Location/Qualifiers
     source          1. .1739
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="W303-1A"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     gene            234. .1523
                     /gene="TFIIIA"
     CDS             234. .1523
                     /gene="TFIIIA"
                     /note="zinc finger containing protein"
                     /codon_start=1
                     /product="transcription factor IIIA"
                     /protein_id="AAB08014.1"
                     /db_xref="GI:172903"
                     /translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
                     RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
                     SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
                     HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
                     HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
                     HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
                     EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
                     ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
ORIGIN

  Query Match             17.5%;  Score 217.4;  DB 8;  Length 1739;
  Best Local Similarity   59.4%;  Pred. No. 1.3e-25;
  Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
              | || || |  ||| | ||    ||||| || |||||||| || | ||   |||| |||
Db        367 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 426

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
              | ||     |||   |  || |     | | | |  |         || |      ||
Db        427 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 483

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
              | ||||  ||| | ||  | | |||    || || || | |||  |     | ||||| |
Db        484 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 543

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
              | |  |||||||||| ||||    ||||| ||||| || |   |||| || ||  |||
Db        544 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 603

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
              |  |||| ||| | || ||||| || || || |  ||| ||   |||   ||  |
Db        604 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 663

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
               ||| || || |||| |   |||||    ||||| ||||||||||||  | | | | |||
Db        664 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 723

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
              | |||  ||| || ||||||    ||      | || |  |   ||  | | || |  |
Db        724 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 783
```

```
Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
              |||||||      ||  | |     | |||| ||||||  |  | |||||||||  |||
Db        784 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 843

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
              | || |  |  ||||||  ||| ||||  |||||||||  |    |||||  ||| | ||||
Db        844 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 903

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
              ||||||  ||| | |||   |||||  ||| |||| ||||||      || ||| | | ||
Db        904 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 963

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
              |||| || ||     | | |||||  ||||  ||| ||   |||  ||    |   |||
Db        964 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1023

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
              | |  ||| ||||  |  |    ||||||
Db       1024 CTAGAAAACATGATCTTCTCACGCATTAT 1052


RESULT 11
YSCP9677/c
LOCUS       YSCP9677               37497 bp    DNA     linear   PLN 01-AUG-1997
DEFINITION  Saccharomyces cerevisiae chromosome XVI cosmid 9677.
ACCESSION   U25841 U00094
VERSION     U25841.1  GI:786295
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 37497)
  AUTHORS   Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Du,Z.,
            Favello,A., Fulton,L., Gattung,S., Greco,T., Kirsten,J., Kucaba,T.,
            Hallsworth,K., Hawkins,J., Hillier,L., Jier,M., Johnson,D.,
            Johnston,L., Langston,Y., Latreille,P., Le,T., Mardis,E.,
            Menezes,S., Miller,N., Nhan,M., Pauley,A., Peluso,D., Rifken,L.,
            Riles,L., Taich,A., Trevaskis,E., Vignati,D., Wilcox,L.,
            Wohldman,P., Vaudin,M., Wilson,R. and Waterston,R.
  TITLE     The sequence of a portion of the right arm of Saccharomyces
            cerevisiae chromosome XVI
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 37497)
  AUTHORS   Miller,N.
  TITLE     The sequence of S. cerevisiae cosmid 9677
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 37497)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1995) Robert Waterston
COMMENT     Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA
                      e-mail: mj@sequencer.wustl.edu

            NEIGHBORING COSMID INFORMATION:
            This sequence includes nucleotides 1-37497 of cosmid 9677.  The
            cosmid on the left is 9705.
FEATURES             Location/Qualifiers
     source          1. .37497
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C (AB972)"
                     /db_xref="taxon:4932"
                     /map="XVI"
                     /note="baker's yeast"
     gene            complement(2366. .3655)
                     /gene="TFC2"
     CDS             complement(2366. .3655)
                     /gene="TFC2"
                     /note="P9677.9"
                     /codon_start=1
                     /product="Transcription factor TFIIIA (PIR accession
                     number S20050)"
                     /protein_id="AAB64615.1"
                     /db_xref="GI:786305"
                     /translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
                     RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
                     SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
                     HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
                     HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
                     HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
                     EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
                     ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
     gene            3890. .4433
                     /gene="RPB6"
     CDS             join(3890. .3909,3986. .4433)
                     /gene="RPB6"
                     /note="P9677.8"
                     /codon_start=1
                     /product="23 kD polypeptide of RNA Polymerases I, II, and
                     III (Swiss Prot. accession number P20435)"
                     /protein_id="AAB64616.1"
                     /db_xref="GI:786304"
                     /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                     GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                     MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
     gene            complement(4621. .5112)
                     /gene="P9677.10"
     CDS             complement(4621. .5112)
                     /gene="P9677.10"
                     /note="Similar to calmodulins (e.g., Swiss Prot. accession
                     number P13868) and calmodulin-related proteins (e.g.,
                     Swiss Prot. accession number P27482) from several
                     organism."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB64617.1"
                     /db_xref="GI:786306"
                     /translation="MDHSESLTFNQLTQDYINKLKDAFQMLDEDEDGLISRGDLTKIY
                     ATLGKTLTDEEWSKMVPDNDTSTAEVGEEGVSFPIFLSIMGKNLSQFPEREELEESLK
                     AIGRGHDLNVPLNEVIDSLKEAGFENPEEEFAKLFKLFTTNQQATEERTFRGKLFLDS
                     ITD"
     gene            5297. .9595
                     /gene="SKI3"
     CDS             5297. .9595
                     /gene="SKI3"
                     /note="P9677.7"
                     /codon_start=1
                     /product="Contains 8 copies of the TPR domain"
                     /protein_id="AAB64618.1"
                     /db_xref="GI:786303"
                     /translation="MSDIKQLLKEAKQELTNRDYEETIEISEKVLKLDPDNYFAHIFL
                     GKALSSLPASNNVSSNRNLERATNHYVSAAKLVPDNLLAWKGLFLLFRTTEVVPDILS
                     YDEYFDLCGQYADALLKQESQSVELINDIKLLKKTHPDCQKAFYQHLKPGSLMAETIG
                     RHLSTPQDALLNLIKILSNIETTEIGKTLSQNRLKLKASDPDYQIKLNSFSWEIIKNS
                     EIDQLYNQLVNILADDQKRSEIENQWLEYRIKVLKSMPLDVKKDFFTKVKEMVEDMVL
                     VNHQSLLAWQKYFEWTDYEDLDNMDAPLIIKYFKKFPKDPLAMILYSWLSSKLSKYDI
                     KSLESANKPPEGHKKTEKETDIKDVDETNEDEVKDRVEDEVKDRVEDEVKDQDEEAKE
                     DEEEDLDDIEIGLLEEEVVTVLTENIVKCKNNILAHRILCQYYLLTKEYEAALPYIKN
                     GISLIAYNIKDLGVHLPLTKREFSLDLATVYTYVDAPKDHNAALKLYDNILSGDFSNI
                     QAKMGKGIIFIERKNWKDAMTLLTQVHEQSPNNLEVLSELSWSKAHMGYMDEALAGLD
                     TVIKGIKGMDLRSIDFRALNLWRQAKVYIMKHASINDAKQENVKCAFKLLIQSIKILD
                     TFAPGFSTLGDIYCHYYKDHLRAFKCYFKAFDLDAGDYTAAKYITETYASKPNWQAAS
                     SIASRLIKGEKAKAELRSNNWPFRVVGIAHLEKQEESDSIEWFQSALRVDPNDVESWV
                     GLGQAYHACGRIEASIKVFDKAIQLRPSHTFAQYFKAISLCDVGEYLESLDILEKVCQ
                     EAATEESFQIGLVEVLMRCSLDLYSQGFLLKSVSIAKDTIERIKIIISELKCENQQVW
                     IYLSQVLRLFIWIESKVDTLPVESLVSIFENSQFSGSEEIDSVDNIKIDTLLDSTTDD
                     NVSIACKFLILASKYSVSDQKFTDIAGTVRASYWYNIGISELTAFITLKEPQYRDAAI
                     FAFKKSIQLQSNTSETWIGLGIATMDINFRVSQHCFIKATALEPKATNTWFNLAMLGL
                     KKKDTEFAQQVLNKLQSLAPQDSSPWLGMALILEEQGDIIGSSKLFAHSFILSNGRSK
                     AAQFMYAKNVLENHINNGDDERDIETVEKLTTASIALEQFFKKSPDSQFALQCALLTL
                     ERLHHYENANELANRLIGILEKKFEKTQDERELFNFAIIKGQFARIHLGLGNFELSIE
```

```
                NADLSQGIISESSDEKSMKTKISNHICLGLSYFFLNDFDQTLNQFQELLSISKDSKHL
                VVLIAKVLYDVGESDTKEIALQELTEYIATSGADLLVTLTIAAMSILDDKREDLSIIL
                EELKALPLSKQIIDKHKDAPYLIEEITKRLYRNDTGKQVWQRSAYFFPNNLKVWERLD
                KNIQRRIASNGQNKVTAEEMSKLYCESKNLRSIQRGMFLCPWNVTAVKALNECF"
gene            complement(9710. .11674)
                /gene="RPC3"
CDS             complement(9710. .11674)
                /gene="RPC3"
                /note="P9677.11"
                /codon_start=1
                /product="74 kD subunit of RNA Polymerase III (Swiss Prot.
                accession number P32349)"
                /protein_id="AAB64619.1"
                /db_xref="GI:786307"
                /translation="MDELLGEALSAENQTGESTVESEKLVTPEDVMTISSLEQRTLNP
                DLFLYKELVKAHLGERAASVIGMLVALGRLSVRELVEKIDGMDVDSVKTTLVSLTQLR
                CVKYLQETAISGKKTTYYYYNEEGIHILLYSGLIIDEIITQMRVNDEEEHKQLVAEIV
                QNVISLGSLTVEDYLSSVTSDSMKYTISSLFVQLCEMGYLIQISKLHYTPIEDLWQFL
                YEKHYKNIPRNSPLSDLKKRSQAKMNAKTDFAKIINKPNELSQILTVDPKTSLRIVKP
                TVSLTINLDRFMKGRRSKQLINLAKTRVGSVTAQVYKIALRLTEQKSPKIRDPLTQTG
                LLQDLEEAKSFQDEAELVEEKTPGLTFNAIDLARHLPAELDLRGSLLSRKPSDNKKRS
                GSNAAASLPSKKLKTEDGFVIPALPAAVSKSLQESGDTQEEDEEEEDLDADTEDPHSA
                SLINSHLKILASSNFPFLNETKPGVYYVPYSKLMPVLKSSVYEYVIASTLGPSAMRLS
                RCIRDNKLVSEKIINSTALMKEKDIRSTLASLIRYNSVEIQEVPRTADRSASRAVFLF
                RCKETHSYNFMRQNLEWNMANLLFKKEKLKQENSTLLKKANRDDVKGRENELLLPSEL
                NQLKMVNERELNVFARLSRLLSLWEVFQMA"
gene            12014. .13120
                /gene="QCR2"
CDS             12014. .13120
                /gene="QCR2"
                /note="P9677.6"
                /codon_start=1
                /product="Ubiquinol-cytochrome C reductase core protein 2"
                /protein_id="AAB64620.1"
                /db_xref="GI:786302"
                /translation="MLSAARLQFAQGSVRRLTVSARDAPTKISTLAVKVHGGSRYATK
                DGVAHLLNRFNFQNTNTRSALKLVRESELLGGTFKSTLDREYITLKATFLKDDLPYYV
                NALADVLYKTAFKPHELTESVLPAARYDYAVAEQCPVKSAEDQLYAITFRKGLGNPLL
                YDGVERVSLQDIKDFADKVYTKENLEVSGENVVEADLKRFVDESLLSTLPAGKSLVSK
                SEPKFFLGEENRVRFIGDSVAAIGIPVNKASLAQYEVLANYLTSALSELSGLISSAKL
                DKFTDGGLFTLFVRDQDSAVVSSNIKKIVADLKKGKDLSPAINYTKLKNAVQNESVSS
                PIELNFDAVKDFKLGKFNYVAVGDVSNLPYLDEL"
gene            14493. .15410
                /gene="P9677.5"
CDS             14493. .15410
                /gene="P9677.5"
                /note="Belongs to the MIP family of transmembrane channel
                proteins."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB64621.1"
                /db_xref="GI:786301"
                /translation="MSSNDSNDTDKQHTRLDPTGVDDAYIPPEQPETKHHRFKISRDT
                LRDHFIAAVGEFCGTFMFLWCAYVICNVANHDVALVAAPDGSHPGQLIMIAIGFGFSV
                MFSIWCFAGVSGGALNPAMSLSLCLARAVSPTRCVVMWVSQIVAGMAAGGAASAMTPG
                EVLFANSLGLGCSRTRGLFLEMFGTAILCLTVLMTAVEKRETNFMAALPIGISLFIAH
                VALTAYTGTGVNPARSLGAAVAARYFPHYHWIYWIGTLLGSILAWSVWQLLQILDYTT
                YVTAEKAASTKEKAQKKGETSSSSAVAEV"
gene            complement(15542. .16012)
                /gene="P9677.12"
CDS             complement(15542. .16012)
                /gene="P9677.12"
                /note="Similar in C-terminus to Gentamicin
                3'-acetyltransferase of P. aeruginosa (Swiss Prot.
                accession number P23181)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB64622.1"
                /db_xref="GI:786308"
                /translation="MSNTSEDNITVRFVTENDKEGWQRLWKSYQDFYEVSFPDDLDDF
                NFGRFLDPNIKMWAAVAVESSSEKIIGMINFFNHMTTWDFKDKIYINDLYVDENSRVK
                GAGGKLIQFVYDEADKLGTPSVYWCTDESNHRAQLLYVKVGYKAPKILYKRKGY"
gene            complement(16937. .19570)
                /gene="P9677.13"
CDS             complement(16937. .19570)
                /gene="P9677.13"
                /note="Similar to S. cerevisiae hypothetical protein
                HRD799 (PIR accession number S45161) and S. pombe

 Query Match           17.5%;  Score 217.4;  DB 8;  Length 37497;
 Best Local Similarity  59.4%;  Pred. No. 6.4e-26;
 Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
               | || || |  ||| | ||    ||||| || |||||||| || | ||   |||| |||
Db       3522 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 3463

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
               | ||     ||||   |  || |     | | | |      || |    ||
Db       3462 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 3406

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
               | ||||  ||| | ||  | | |||     || || || | |||  |      | |||||
Db       3405 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 3346

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
               | |  |||||||||||| |||||     ||||| ||||| || |   |||| || ||  |||
Db       3345 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 3286

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
               |  |||| ||| | || ||||| || || || |  ||| ||   |||   ||  |
Db       3285 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 3226

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
               ||| || || |||| |   |||||     ||||| ||||||||||||  | | | | |||
Db       3225 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 3166

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
               | |||  ||| || ||||||   ||     | || |  |   ||  | | || |  |  |
Db       3165 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 3106

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
               |||||||      ||  | |      | |||| ||||||  |  | |||||||||  |||
Db       3105 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 3046

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
               | || |  |  ||||||  ||| ||||  |||||||||  |    |||||  ||| | ||||
Db       3045 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 2986

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
               ||||||  ||| | |||   ||||| ||| |||| ||||||      || ||| | | ||
Db       2985 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 2926

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
               |||| || ||      | | |||||  |||| ||| ||   |||  ||    |   |||
Db       2925 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 2866

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
               | |  ||| |||  |  |    ||||||
Db       2865 CTAGAAAACATGATCTTCTCACGCATTAT 2837


RESULT 12
YSCRP026A/c
LOCUS       YSCRP026A              2133 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  S.cerevisiae RNA polymerase II sixth subunit (RP026) gene, complete
            cds.
ACCESSION   M33924
VERSION     M33924.1  GI:172452
KEYWORDS    RNA polymerase II; c-myc proto-oncogene; transmembrane protein;
            tyrosine kinase.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2133)
```

```
  AUTHORS   Archambault,J., Schappert,K.T. and Friesen,J.D.
  TITLE     A suppressor of an RNA polymerase II mutation of Saccharomyces
            cerevisiae encodes a subunit common to RNA polymerases I, II, and
            III
  JOURNAL   Mol. Cell. Biol. 10 (12), 6123-6131 (1990)
  MEDLINE   91061718
   PUBMED   2247052
COMMENT     Original source text: S.cerevisiae single-copy plasmid DNA, clone
            2.
            Draft entry and computer-readable sequence for [Unpublished (1990)]
            kindly submitted
            by J.D.Friesen, 01-MAY-1990.
               The Hospital for Sick Children
               555 University Avenue
               Toronto, Ontario, CANADA M5G 1X8
               e-mail: jim@sickkids.toronto.edu.
FEATURES             Location/Qualifiers
     source          1. .2133
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     CDS             join(1026. .1045,1122. .1569)
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /codon_start=1
                     /protein_id="AAA34989.1"
                     /db_xref="GI:172453"
                     /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                     GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                     MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
     exon            <1026. .1045
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=1
     intron          1046. .1121
                     /note="RP026 intron A"
     exon            1122. .>1569
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=2
ORIGIN      Chromosome XVI.

  Query Match             16.1%;  Score 199.2;  DB 8;  Length 2133;
  Best Local Similarity   61.4%;  Pred. No. 1.1e-22;
  Matches  340;  Conservative    0;  Mismatches  208;  Indels    6;  Gaps    1;

Qy        171 TACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGT 230
              | |||||    | ||||  ||| | ||  | | |||    || || || | |||  |
Db        556 TTCAGTGTGATAGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTA 497

Qy        231 ATCACATTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCG 290
              | ||||| || |  |||||||||| |||||    ||||| ||||| || |    ||||
Db        496 TACGCATTCTGATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCG 437

Qy        291 ACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAA 350
              || ||  ||||  |||| ||| | || ||||| || || || |  ||| ||   |||
Db        436 CCAGCAACTGAAGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGG 377

Qy        351 TTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410
              ||  |      ||| || || |||| |   |||||    ||||| ||||||||||||
Db        376 ATGCAACCTCCGATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTT 317

Qy        411 AAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACA 470
              | | | | |||| |||  ||| || ||||||   ||      | || |  |   ||  | |
Db        316 ACATAAACTAACCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAA 257

Qy        471 ACATAAATTAAAACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGG 524
              || |  |  ||||||||      ||  | |     | |||| ||||||  |  | ||||
Db        256 TCACATTTCTAAACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGG 197

Qy        525 TTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCC 584
              |||||  ||| | || |  | ||||||  ||| |||| |||||||||  |    |||||
Db        196 TTGTTGCAAAGAGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCC 137

Qy        585 AAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACA 644
              ||| | ||||||||||  ||| | |||   ||||| ||| |||| ||||||      ||

Db        136 TAAATTAAAATGTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCA 77

Qy        645 TATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGT 704
              ||| | | |||||| || ||     | | |||||  ||||  ||| ||   |||  ||
Db         76 CATGATTATTCATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGA 17

Qy        705 GGGGAAATTTGCAA 718
                |   ||| | |
Db         16 TATGTCTTTTTCTA 3


RESULT 13
CNS06ING
LOCUS       CNS06ING                 797 bp    DNA     linear   STS 30-NOV-2001
DEFINITION  T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL400514
VERSION     AL400514.1  GI:12156635
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 797)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
  MEDLINE   20584715
   PUBMED   11152880
REFERENCE   3  (bases 1 to 797)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .797
                     /organism="Saccharomyces bayanus"
                     /mol_type="genomic DNA"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA022B07"
                     /clone_lib="AS0AA"
                     /note="end : T7"
     misc_feature    <2. .>796
                     /note="similar to Saccharomyces cerevisiae ORF YPR186c [
                     PZF1 ; TFIIIA (transcription initiation factor) ]"
                     /evidence=not_experimental
```

```
ORIGIN

  Query Match             15.6%;  Score 193;  DB 11;  Length 797;
  Best Local Similarity   59.3%;  Pred. No. 1.4e-21;
  Matches  347;  Conservative    1;  Mismatches  231;  Indels    6;  Gaps    1;

Qy        178 GACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACAT 237
              ||| ||||||  ||| | ||| | | :||    ||| | || |  ||  |  | || |||
Db         83 GACAATTGTGCAAAATCCTTTGTTAARAAGAGTCATCTAGAGAGGCACTTGTTTTCGCAT 142

Qy        238 TCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACAC 297
              || || |  || |||||||| |||||    |||||| || |||||  |   |||||| ||
Db        143 TCTGATACGAAGCCATTCCAATGTTCTTATTGTGGCAAGGGGGTGACAACTCGACAGCAA 202

Qy        298 TTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAA 357
              |||||  | || ||| | || ||||| || || |||   ||| |    ||    ||| |
Db        203 TTGAAGCGCCACGAAGTTACACATACCAAATCTTTTGTTTGTCCGGAAGATGGATGTGAT 262

Qy        358 GAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACA 417
              |||||||||| || |      |||||   ||||||  ||||||| ||    | | | |
Db        263 CTCCGATTTTATAAGCACCCTCAATTAAGGGCTCATATTCTATCTGTCCACCTAGACAAA 322

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              ||||| |||  |||||||| ||||  |||     |||| |  |   ||    | ||||
Db        323 TTAACCTGTCCACAATGTAGTAAAARTTTTCAGAGACCCTACAGGCTAAGGAATCATATC 382

Qy        478 TTAAAACATCATGGT------GGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTT 531
              |  |||||||||| |      |       | | |||||||||  |    |||| |||| |
Db        383 TCTAAACATCATGATCCTGAAGTGGTAAATCCGTATCAATGTACTTTCGCTGGCTGTTGT 442

Qy        532 AAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTT 591
              | | | | ||   ||||||   || ||||  ||||| |||  |     |||| ||| |
Db        443 ACAGAGTACCGTATATGGTCACAGTTGCAATCACATATCAAAAATGACCATCCTAAATTG 502

Qy        592 AAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTA 651
              ||||| || |  ||    ||  |||||| ||| |||| ||| ||     ||||||
Db        503 AAATGCCCCATTTGCAACAAGCCTTGTGTAGGGGAAAATGGTCTACAAATGCATATGATC 562

Qy        652 AGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAA 711
              | |||||||||| |||    | |  || |  |||| ||| ||  ||||  |
Db        563 ATTCATGATGAATCTTTGGTAACTAAGAACTGGAAATGTCATATTTGTTCTAATACATCT 622

Qy        712 TTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCCAT 756
              ||    |  || | ||  |  ||   || ||  | ||    |||
Db        623 TTCTTGAGAAAGCACGACCTCCTTACTCACTACCAGATGACGCAT 667


RESULT 14
AE016818_01
WPCOMMENT
  Sequence split into 16 fragments  LOCUS AE016818 Accession AE016818
     Fragment Name      Begin     End
     AE016818_00            1    110000
     AE016818_01       100001    210000
     AE016818_02       200001    310000
     AE016818_03       300001    410000
     AE016818_04       400001    510000
     AE016818_05       500001    610000
     AE016818_06       600001    710000
     AE016818_07       700001    810000
     AE016818_08       800001    910000
     AE016818_09       900001   1010000
     AE016818_10      1000001   1110000
     AE016818_11      1100001   1210000
     AE016818_12      1200001   1310000
     AE016818_13      1300001   1410000
     AE016818_14      1400001   1510000
     AE016818_15      1500001   1519138
Continuation (2 of 16) of AE016818 from base 100001 (AE016818 Eremothecium gossypii ATCC

  Query Match              9.7%;  Score 120.6;  DB 8;  Length 110000;
  Best Local Similarity   50.1%;  Pred. No. 1.9e-10;
  Matches  386;  Conservative    0;  Mismatches  374;  Indels   11;  Gaps    3;

Qy         25 TCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAAAGTATATTTGCACA 84
              ||| | ||   | |    ||  | | ||      | || | ||||| | ||   |||
Db      17349 TCGTTGTCGAGTGTTGAGTCCACGGCGTCGAGCACGCGGCGCAAAACGCATTACTGCGAC 17408

Qy         85 TATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACA-TTTAAGAAC 143
              || || || ||  | || || |  |  || || ||  |  |   | | ||     |||
Db      17409 TACGAGGGCTGCTACAAGGCGTTCACGCGGCCGTCCCTGCTGACGGAGCACCAGCAGACG 17468

Qy        144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAG 203
              | || |  |     | |||    || |    || | ||| |       |  |||  ||
Db      17469 GCGCACCAGGGCATCAGGGCATACCAGTGC----GAGCAGTGCGGGCGCGGGTTTACCAA 17524

Qy        204 AAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTC 263
              || || || |||||    ||  |  | || || || || | ||| || |||    ||  |
Db      17525 GAAGTCGCACTTGGAGCGGCACCTGTTCTCGCACTCTGAGACAAAGCCGTTCAGCTGCAC 17584

Qy        264 AGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATAC 323
              || ||||| || ||||| |   | || || ||  |||   | || || ||||| || ||
Db      17585 GGTATGTGGCAAGGGGGTGACCACGCGGCAGCAACTGAGGCGGCACGAGATCACACACAC 17644

Qy        324 AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTT 383
              |||||| || || ||    |     ||    ||    ||||| || || || || |     ||
Db      17645 AAAGTCGTTCAAGTGCCCGCACGAGGGCTGCGGCGAAGCGTTCTACAAGCACCCGCAGTT 17704

Qy        384 AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGT 443
              |    || |  |  | || || ||  |||   | ||||| | ||| ||| | ||
Db      17705 GCGTTCGCACGTGCTGGCGGTGCACGAGCAAAAGCTGACGTGCACACACTGTGACAAGCG 17764

Qy        444 TTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCC--- 500
              |||   || || |     ||    | ||| |     || ||||| ||      |  |
Db      17765 GTTCCAACGGCCATACCGTTTGAAAACACACATCGCCAAGCATCACGGGCCTGCATCGCA 17824

Qy        501 ---TGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATT 557
                  || |||||     |  | || ||  |     ||  | || ||||| |    |
Db      17825 GTTCCGGTACCAATGCACGAACGCCGGCTGCGTACAGTGCTTTGAGACGTGGTCTGCGCT 17884

Qy        558 ACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTG 617
              ||    ||  |  |          || || || ||  |||        || || ||    ||
Db      17885 GCAGCAGCACCTTCACACGGACCACCCGAAGCTACCATGCGGGGTGTGCGGGAAGCTGTG 17944

Qy        618 TGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAA 677
              || ||| |  |  ||  |        || |||      || || || ||     |||| ||
Db      17945 CGTGGGGGAGACGGGACTGCAGATGCACATGACTGTGCACGACGAGTCGCGGGTGATTAA 18004

Qy        678 AATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGA 737
              ||  |||| ||   |  |||  |      |  |  || ||||     ||  |  |||
Db      18005 AAACTGGAAGTGCTCTGTGTGTAGCGACACCACCTACGCGAAGATGGCCGATCTGCTTGC 18064

Qy        738 ACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGA 788
              |||||||  |   ||| | || |  |||||  | || ||  || || |
Db      18065 GCATTATATGAAAACCCACGGTGATTCGATCCCCAAGGAATTGATAGAGCA 18115


RESULT 15
AC138126/c
LOCUS       AC138126              169063 bp    DNA     linear   PRI 17-DEC-2002
DEFINITION  Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
ACCESSION   AC138126
VERSION     AC138126.1  GI:27151357
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169063)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169063)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
```

```
            Drive, Walnut Creek, CA 94598, USA
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .169063
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-274A19"
ORIGIN

  Query Match             8.2%;  Score 101;  DB 9;  Length 169063;
  Best Local Similarity   50.8%;  Pred. No. 2.4e-07;
  Matches  298;  Conservative    0;  Mismatches  280;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |           || ||  |  |  || |    |      |   | | |
Db      80134 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 80075

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
             | || |    | || || |||||  ||||  |||||  |   |||||  | |||  | |
Db      80074 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 80015

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||||   | ||  || | |   ||   ||| || | ||||      || || |||
Db      80014 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 79961

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |      ||||| | |  |||||||  |||   ||| | | | |
Db      79960 GGCAAAGCCTTTAACCATTCCTCACATCTTGCTACACATAAGGTAATTCATACTGGAGAG 79901

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  ||||||||||   ||||        |||| | |  |||
Db      79900 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 79841

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  | |||||               |    || |||||| |||    |||| |||
Db      79840 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 79781

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              | |      |    | |   ||||| | | | |   | |  ||| |||   |  |
Db      79780 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 79721

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  ||||||  |  | |||        |||||| ||  |  ||||||||  |||
Db      79720 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 79661

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||      ||     |       ||   |          ||||||| ||  ||
Db      79660 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 79601

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              ||||| |     ||  ||    |||  | ||    | | | || |
Db      79600 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTACAAGT 79554


Search completed: February  9, 2005, 05:36:23
Job time : 5295.91 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 22:43:04 ; Search time 642.278 Seconds
(without alignments)
11419.610 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1239	100.0	2060	3	AAA15398	Aaa15398 DNA encod
	2	1235.8	99.7	1239	6	ABZ32206	Abz32206 Candida a
	3	99.4	8.0	2132	10	ADA53158	Ada53158 Human cod
	4	96.8	7.8	1705	9	ACA98970	Aca98970 cDNA enco
	5	96.6	7.8	2320	10	ACA56456	Aca56456 Human sig
	6	96.6	7.8	2320	12	ADI56252	Adi56252 Human pol
	7	95.2	7.7	4563	5	AAS91317	Aas91317 DNA encod
	8	95	7.7	981	12	ADN04717	Adn04717 Antipsori
	9	94.6	7.6	2597	8	AAL51569	Aal51569 Human nuc
	10	92.4	7.5	3639	5	AAS64586	Aas64586 DNA encod
c	11	92	7.4	994	12	ACH87568	Ach87568 Human gen
	12	92	7.4	2873	13	ADR14368	Adr14368 Human NF-
	13	91.8	7.4	1533	12	ADN99064	Adn99064 Novel hum
	14	91.8	7.4	1533	12	ADO00633	Ado00633 Novel hum
	15	91.8	7.4	2110	10	ADB63624	Adb63624 Human cDN
	16	91.8	7.4	2760	10	ADC56695	Adc56695 Human mac
	17	91.8	7.4	3078	10	ADB62468	Adb62468 Human cDN
	18	91	7.3	3755	13	ADR07315	Adr07315 Full leng
	19	91	7.3	4209	13	ADR08223	Adr08223 Full leng
	20	90.4	7.3	2026	8	AAD55863	Aad55863 Human nuc
	21	90.4	7.3	2114	10	ADA53124	Ada53124 Human cod
	22	90.2	7.3	2230	10	ADA52931	Ada52931 Human cod
	23	89.8	7.2	590	12	ACH80103	Ach80103 Human gen
	24	89.2	7.2	2064	10	ADC30762	Adc30762 Human nov
	25	89.2	7.2	2221	11	ADM01344	Adm01344 Human cDN
	26	89.2	7.2	2622	5	AAS68872	Aas68872 DNA encod
	27	89.2	7.2	2729	4	AAH16178	Aah16178 Human cDN
	28	89.2	7.2	3839	6	ABK83826	Abk83826 Human cDN
	29	89.2	7.2	3839	12	ADN04207	Adn04207 Antipsori
	30	89.2	7.2	3839	13	ADR25049	Adr25049 Breast ca
	31	89	7.2	1991	13	ADS09824	Ads09824 Human the
	32	89	7.2	2298	4	AAH16608	Aah16608 Human cDN
	33	89	7.2	2299	13	ADS09825	Ads09825 Human the
	34	89	7.2	2905	5	AAS66143	Aas66143 DNA encod
	35	89	7.2	3003	12	ADN06033	Adn06033 Antipsori
	36	89	7.2	3020	5	AAS92560	Aas92560 DNA encod
	37	89	7.2	3502	4	AAI57845	Aai57845 Human pol
	38	88.6	7.2	1335	12	ADN98906	Adn98906 Novel hum
	39	88.6	7.2	1335	12	ADO00475	Ado00475 Novel hum
	40	88.6	7.2	1466	12	ACH91879	Ach91879 Human gen
	41	88.6	7.2	1875	10	ADI60591	Adi60591 Secreted
	42	88.6	7.2	2597	8	AAD55855	Aad55855 Human nuc
	43	88.6	7.2	4227	9	ACA98938	Aca98938 cDNA enco
c	44	88.4	7.1	1410	12	ACH87337	Ach87337 Human gen
	45	88.4	7.1	1952	10	ADC58104	Adc58104 Zinc fing

ALIGNMENTS

RESULT 1
AAA15398
ID AAA15398 standard; DNA; 2060 BP.
XX
AC AAA15398;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding a transcription factor designated CATFIIIA.
XX
KW Transcription factor; CATFIIIA; DNA-binding protein;
KW ribosomal RNA 5S gene; fungal infection; ss.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT CDS 720. .1958
FT /*tag= a
FT /transl_except= (pos: 1296. .1298, aa: Ser)
FT /transl_except= (pos: 1734. .1736, aa: Ser)
XX
PN WO200028037-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002739.
XX
PR 10-NOV-1998; 98FR-00014147.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Camier S, Sentenac A;
XX
DR WPI; 2000-376549/32.
DR P-PSDB; AAY93316.
XX
PT New nucleic acid encoding Candida albicans transcription factor, useful
PT e.g. in screening for antimycotic agents and for immunization.
XX
PS Claim 4; Page 32-33; 45pp; French.
XX
CC The present sequence encodes a Candida albicans transcription factor,
CC designated CATFIIIA. The polypeptide is a DNA-binding protein, which is

```
CC   involved in initiating transcription of the ribosomal RNA 5S gene. The
CC   polynucleotide is used to screen for its specific inhibitors, potentially
CC   useful as antimycotic agents, to raise an antibody response that is
CC   protective against fungal infection and to raise antibodies. Such
CC   antibodies, as well as the polypeptides and polynucleotides are used in
CC   compositions for diagnosing and treating fungal infections, e.g. by
CC   detecting polymorphisms and mutations
XX
SQ   Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1239;  DB 3;  Length 2060;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-234;
  Matches 1239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1958


RESULT 2
ABZ32206
ID   ABZ32206 standard; DNA; 1239 BP.
XX
AC   ABZ32206;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential gene SEQ ID NO 6493.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   P-PSDB; ABP73656.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 37; SEQ ID NO 6493; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
```

```
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans gene used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office
XX
SQ   Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

  Query Match             99.7%;  Score 1235.8;  DB 6;  Length 1239;
  Best Local Similarity   99.8%;  Pred. No. 1.4e-233;
  Matches 1237;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              ||||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239


RESULT 3
ADA53158
ID   ADA53158 standard; cDNA; 2132 BP.
XX
AC   ADA53158;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 726.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54797.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
```

PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 726; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

Query Match 8.0%; Score 99.4; DB 10; Length 2132;
Best Local Similarity 50.6%; Pred. No. 4.7e-10;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |         || ||  |  |  || |   |      |   | | |
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |    | || || |||||  ||||  |||||  |   |||||  | |||  | |
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||   | ||  || | |   ||  ||| || | ||||      || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |       |||||  | |  ||||||| |||   ||| | | | |
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  ||||||||||   ||||          |||| | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  |  ||||        |    || |||||| |||    |||| |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |      |    | |   ||||| | | | |   |  |  ||| |||   | |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||         |||||| ||  |  ||||||||  |||
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||      ||   |      ||   |        ||||||| ||  ||
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              |||||  |      ||  ||     |||  | ||    | | | || |
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964
```

RESULT 4
ACA98970
ID ACA98970 standard; cDNA; 1705 BP.
XX
AC ACA98970;
XX
DT 25-JUL-2003 (first entry)
XX
DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX
KW Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW cell proliferative disorder; cancer; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; stroke;
KW immune disorder; inflammatory disorder; AIDS; allergy;
KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW protein-protein interaction; drug-target interaction;
KW gene expression profile; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003023003-A2.
XX
PD 20-MAR-2003.
XX
PF 05-SEP-2002; 2002WO-US028540.
XX
PR 07-SEP-2001; 2001US-0317792P.
PR 07-SEP-2001; 2001US-0317912P.
PR 14-SEP-2001; 2001US-0322270P.
PR 21-SEP-2001; 2001US-0324040P.
PR 28-SEP-2001; 2001US-0326732P.
PR 19-OCT-2001; 2001US-0346716P.
PR 25-JAN-2002; 2002US-0351749P.
PR 22-FEB-2002; 2002US-0359498P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ;
PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BM;
PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI Hafalia AJA, Sanjanwala B, Marquis JP, Gorvad AE, Lee SY, Ison CH;
PI Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebarjadian Y, Shah P;
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE;
PI Burford N, Ramkumar J;
XX
DR WPI; 2003-313243/30.
DR P-PSDB; ABU96722.
XX
PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
XX
PS Claim 5; Page 340-341; 345pp; English.
XX
CC The invention describes a novel human isolated nucleic acid-associated
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. These are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The microarray is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles.
CC This sequence encodes a novel human nucleic acid-associated protein
CC (NAAP)
XX
SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.8; DB 9; Length 1705;
Best Local Similarity 52.9%; Pred. No. 1.5e-09;
Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||  || ||     |    |   |   |
Db        695 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 754

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              |  | || | | || || |||||  | ||  | ||||| |||  |  | |||    |
Db        755 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 814
```

```
Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  |||||   ||||  || | |   ||   ||| || |||||||  |      || |||
Db        815 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 868

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              | ||||| |||  |   |  |||||  |    | |||||    |   ||| | ||| |
Db        869 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 928

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  | | | |||  ||  ||||| ||||   ||||      |    | | |  | |
Db        929 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 988

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||    |  | ||   |  | || ||| || |||     ||| |||
Db        989 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1048

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
              |  ||    |   | |   | ||||  |||  | |     || |||   | |
Db       1049 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAATGCTACAAA 1108

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||||||| |  ||| |||||||  |  || | ||||||  ||
Db       1109 TGTGAAGAATGTGGCAAAGTTTTTAACCGATCTTCAAACCTTACAGACCATAAAAAAATT 1168

Qy        484 CATCATGGTG 493
              |||   || |
Db       1169 CATACAGGAG 1178
```

RESULT 5
ACA56456
ID ACA56456 standard; cDNA; 2320 BP.
XX
AC ACA56456;
XX
DT 06-JUN-2003 (first entry)
XX
DE Human signalling pathway polynucleotide probe SEQ ID NO 1054.
XX
KW Human; probe; ss; array element; Parkinson's disease;
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS Homo sapiens.
XX
PN US6500938-B1.
XX
PD 31-DEC-2002.
XX
PF 30-JAN-1998; 98US-00016434.
XX
PR 30-JAN-1998; 98US-00016434.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Seilhamer JJ;
XX
DR WPI; 2003-352189/33.
XX
PT Combination of polynucleotide probes, useful as array elements in a
PT microarray for monitoring the expression of a number of target
PT polynucleotides.
XX
PS Claim 1; SEQ ID NO 1054; 65pp; English.
XX
CC The invention relates to a combination which, comprises a number of
CC polynucleotide probes comprising a sequence selected from one of the 1490
CC sequences mentioned in the specification. The combination is useful as an
CC array element in a microarray for monitoring the expression of a number
CC of target polynucleotides. The microarray is particularly useful in the
CC diagnosis and treatment of cancer and immunopathology and neuropathology.
CC The microarray is useful in diagnostics and treatment regimens, drug
CC discovery and development, toxicological and carcinogenicity studies,
CC forensics and pharmacogenomics. The microarray is also useful for
CC monitoring progression of diseases and for developing sophisticated
CC profiles for the effects of currently available therapeutic drugs. The
CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC and genomic fragments and in research and diagnostic applications. The
CC array can detect changes in expression in a large number of genes coding
CC for different signaling pathway populations which can be used to diagnose
CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC and Parkinson's disease. The present sequence represents a polynucleotide
CC probe of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=06500938B1
XX
SQ Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

```
  Query Match             7.8%;  Score 96.6;  DB 10;  Length 2320;
  Best Local Similarity   51.3%;  Pred. No. 1.7e-09;
  Matches  306;  Conservative    0;  Mismatches  279;  Indels   12;  Gaps    3;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || | |||  |||| |||||| | ||| ||| | |||     |
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                 |||  || ||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               ||||| |||          ||| |  |    |||||||   ||   ||| | | | | ||
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  |   ||||| ||||  ||||          |||| | |  | |||
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              |||  | |  |||||          |  | || |||||| |||  | |||| ||| |
Db       1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy        369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | |       ||   |  | |||| | ||| |   |  |   || ||| | |  |  ||
Db       1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              |||| |||||  ||||| ||| |   |    ||||||    |   |  |||||||  |||  ||
Db       1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy        486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
              |  ||| |     |||    |        ||   |        ||||||  | |  |  ||
Db       1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy        546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
                 | | |    |       |   |    |  | | ||  |   | |||||   |||||
Db       1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy        606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||| |||| || |     |    | | ||  |||||  |  |||||||   ||
Db       1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551
```

RESULT 6
ADI56252
ID ADI56252 standard; DNA; 2320 BP.
XX
AC ADI56252;
XX
DT 22-APR-2004 (first entry)
XX
DE Human polynucleotide probe #1054.
XX
KW Human; probe; ss; receptor-like polypeptide; transducing polypeptide;

KW effector-like polypeptide; cancer; immunopathology; neuropathology;
KW drug development; toxicology; carcinogenicity;
KW signalling pathway polypeptide; adrenal gland; bladder; bone;
KW bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;
KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;
KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.
XX
OS Homo sapiens.
XX
PN US2004010136-A1.
XX
PD 15-JAN-2004.
XX
PF 26-NOV-2002; 2002US-00305720.
XX
PR 30-JAN-1998; 98US-00016434.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Seilhamer JJ;
XX
DR WPI; 2004-090520/09.
XX
PT New composition comprising polynucleotide probes, useful as array
PT elements in a microarray for monitoring the expression of target
PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic
PT fragments.
XX
PS Claim 6; SEQ ID NO 1054; 73pp; English.
XX
CC The invention relates to a composition of polynucleotide probes
CC comprising first polynucleotide probes comprising at least a portion of a
CC gene encoding a receptor-like polypeptide, second polynucleotide probes
CC comprising at least a portion of a gene encoding a transducing
CC polypeptide and third polynucleotide probes comprising at least a portion
CC of a gene encoding an effector-like polypeptide. The probes of the
CC composition are useful as array elements in a microarray for monitoring
CC the expression of target polynucleotides. The microarray is useful in the
CC diagnosis and treatment of cancer, an immunopathology or a
CC neuropathology. It can also be used for drug discovery and development,
CC toxicological and carcinogenicity studies, forensics or pharmacogenomics.
CC Microarrays can also be used for monitoring the progression of diseases
CC that may be associated with the altered expression of signalling pathway
CC polypeptides. The composition can also be used to purify a subpopulation
CC of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile
CC is also useful for the diagnosis and treatment of cancer, e.g. cancers of
CC the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,
CC an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or
CC ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,
CC epilepsy, Alzheimer's disease or depression. This sequence represents a
CC human polynucleotide probe of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.6; DB 12; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.7e-09;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

```
Qy        69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
             | || |    | || || | |||  |||| ||||||  | ||| ||| | |||    |
Db       964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy       129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||  || ||  || | |   ||  ||| || ||||||      || || ||||
Db      1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy       189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              ||||| |||         ||| |  |    |||||||   ||   ||| | | | | ||
Db      1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy       249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
             ||| | | | |||  |  |||| ||||  ||||        |||| | |  | |||
Db      1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy       309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
             |||  | |  |||||               | || |||||| |||  | |||| ||| |
Db      1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy       369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
             | |       ||   |   | |||| | ||| |   |  |   || ||| | |  |  ||
Db      1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy       426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
             |||| |||||  ||||| ||| |  |  ||||||   |  |  ||||||| ||| ||
Db      1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy       486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
             |  ||| |     |||    |       ||   |       ||||||  | | |  ||
Db      1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy       546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
             | | |    |      |   |   | | ||  |   | |||||   |||||
Db      1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy       606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
             ||| |||| || |     |   | | ||  ||||| |  |||||||   ||
Db      1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551
```

RESULT 7
AAS91317
ID AAS91317 standard; cDNA; 4563 BP.
XX
AC AAS91317;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27121.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27130.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27121; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal

```
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 95.2;  DB 5;  Length 4563;
  Best Local Similarity   52.7%;  Pred. No. 3.5e-09;
  Matches  258;  Conservative    0;  Mismatches  223;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||          || ||      |  ||      |    ||   | | |
Db       1315 GAATGTGGCAAAGCCTTTAACCAGTCCTCACACCTTACTCAACATAAAACAATTCATACT 1374

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | | | |  | || || |||||  ||||  ||||| | ||| |  |  |||    |
Db       1375 GGAGGGAAAACCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATCTTTCAGCCCTT 1434

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||   | ||  || | |   ||   ||| || ||||||      || || |||
Db       1435 ACTAAACATAAGATAATTCATACTGGGGAGAAACCATACAAATGT------GAAGAATGT 1488

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||     |   |||||  |    | |||||  | |   ||| | | | |
Db       1489 GGCAAAGCTTTTAGGCAATCCTCACACCTTACTAGACATAAAGCAATTCATACTGGAGAG 1548

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  ||||| ||||   ||||   |     | || | | ||||
Db       1549 AAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATTTCTCAGACCTTAGAAGA 1608

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||                | | || |||| | |||     |||| |||
Db       1609 CATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGGAAAGCTTTT 1668

Qy        367 TATAAACATCAATCTTTAAGA---CATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                |   | | | |||   ||||| ||| |  |  |  || ||| | | |  |
Db       1669 AGCCAGTCCTCAACCCTTAGAAACCATCAGATAATTCATACTGGAGAGAAACCCTACAAA 1728

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||  ||||| ||| |    |  | |||||| |  |   ||||||  |||
Db       1729 TGTGAAGAATGTGGTAAAGCTTTTAAGTGGTCATCAAAACTTACTGTACATAAGGTAATT 1788

Qy        484 CATCATGGTG 493
              |||  ||| |
Db       1789 CATACTGGAG 1798


RESULT 8
ADN04717
ID   ADN04717 standard; cDNA; 981 BP.
XX
AC   ADN04717;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Antipsoriatic cDNA sequence #572.
XX
KW   ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004028479-A2.
XX
PD   08-APR-2004.
XX
PF   25-SEP-2003; 2003WO-US030907.
XX
PR   25-SEP-2002; 2002US-0414006P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Bodary S,  Clark H,  Jackman J,  Schoenfeld J,  Williams PM,  Wood WI;
PI   Wu TD;
XX
DR   WPI; 2004-305105/28.
DR   P-PSDB; ADN04718.
XX
PT   New PRO nucleic acid or polypeptide, useful for preparing a
PT   pharmaceutical composition for diagnosing or treating psoriasis in a
PT   mammal.
XX
PS   Claim 1; SEQ ID NO 1111; 3069pp; English.
XX
CC   The invention relates to novel polynucleotide and polypeptides for
CC   treating psoriasis or a sequence having at least 80% identity to the
CC   above sequences. The nucleic acid is useful for preparing a composition
CC   for diagnosing or treating psoriasis in a mammal. This sequence
CC   corresponds to one of the polynucleotides of the invention.
XX
SQ   Sequence 981 BP; 380 A; 178 C; 170 G; 253 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 95;  DB 12;  Length 981;
  Best Local Similarity   50.4%;  Pred. No. 3e-09;
  Matches  289;  Conservative    0;  Mismatches  275;  Indels    9;  Gaps    2;

Qy         89 AAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACA 148
              |||  ||||  |||| || || |  |   |||    |     |||||   | ||  || |
Db          1 AAGAATGTGGCAAAGGCTTTAGTATATTCTCAACCCTTACTAAACATAAGATAATTCATA 60

Qy        149 GTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAAT 208
                |   ||   ||| || ||||| |      | || |||| |||||| |||  | |    |
Db         61 CTGGAGAGAAACCCTACAAATGCA------ATGAATGTGGTAAAGCCTTTAACTGGTCCT 114

Qy        209 CACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGT 268
              ||  | |    | ||||     |   ||| | | | | ||||| | | | |||  ||  |
Db        115 CAACTCTTACTAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAAT 174

Qy        269 GTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGT 328
              |||| ||||   ||||          |  | | |  |||||| | ||    |||||
Db        175 GTGGCAAAGCTTTTAACCAGTCCTCAACCCTTACTAGACATAAGATAGTTCATACTGGAG 234

Qy        329 CATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGAC 388
                  |  | || |||| | |||    ||||  ||| |          | || ||| ||
Db        235 AGAAACCCTACAAATGTGAAGAATGTGGTAAAGCCTTTAAACGGTCCACAACTCTAAAAC 294

Qy        389 ATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCA 448
              || | | | | | | | || ||| |||  |  ||| || |||||   ||||  || |
Db        295 ATAAGAGAATTTATACTAAAGAGAAACCATACAAATGTGAAGAATGTGGAAAAGCCTTTA 354

Qy        449 CTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTTATC 508
              |  |   ||||   |  |  |||||||  |||  |||  ||| | |   |||    |
Db        355 GTGTATTCTCAACCCTTACTAAACATAAGATAATTCATACTGGAGCAAAACCTTACAAAT 414

Qy        509 AATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATA 568
                ||   |   ||  ||||     |||  |||     | | |    |       |
Db        415 GTGACGAATGTGGCAGTGCCTTTAGGGCATTCTCAACCCTTACTGAACATAAGAGAGTTC 474

Qy        569 TAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAA 628
                |   | | | ||  |   |||||   | ||||||||||||  | |    ||
Db        475 ATACTGGAGAGAAACCTTAC---AAATGCAATGAATGTGGTAAAGCCTTTAACTGGTCCT 531

Qy        629 AAGGTTTATCTTCACATATGTTAAGTCATGATG 661
                | | |  ||  ||||| |  || ||||  ||
Db        532 CAACTCTTACTAAACATAAGAGAATTCATACTG 564
```

RESULT 9
AAL51569
ID AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID No 52.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
PR 29-JUN-2001; 2001US-0301787P.
PR 29-JUN-2001; 2001US-0301792P.
PR 29-JUN-2001; 2001US-0301892P.
PR 29-JUN-2001; 2001US-0301893P.
PR 06-JUL-2001; 2001US-0303405P.
PR 06-JUL-2001; 2001US-0303442P.
PR 15-MAR-2002; 2002US-0364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowsky ML, Yao MG, Walia NK, Bandman O, Lal PG, Becha SD, Lee SY;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y, Lu Y;
XX
DR WPI; 2003-201420/19.
DR P-PSDB; AAO16419.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS).
XX
PS Claim 12; Page 292-293; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present DNA sequence encodes a human nucleic acid-associated
CC protein of the invention
XX
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

Query Match 7.6%; Score 94.6; DB 8; Length 2597;
Best Local Similarity 55.1%; Pred. No. 4.2e-09;
Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;

```
Qy        78 TTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTT 137
             ||  | || |||||  ||||  ||||||| ||||    | |||    |   ||||||
Db      1070 TTATAAATGTGAAGAATGTGGCAAAGCCTTTAATGTGTCCTCAACCCTTACTCAACATAA 1129

Qy       138 AAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATT 197
              ||||  || | |   ||   ||| || ||||||      || || ||||  ||||| ||
Db      1130 GAGAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAGTGTGGCAAAGCCTT 1183

Qy       198 TTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCA 257
             |  |       |||  | |       |||||     |   ||| | | | |||||||||| | |
Db      1184 TAACGTGTCCTCAACTCTTACTCAACATAAGAGAATTCATACTGGAGAAAAACCATACAA 1243

Qy       258 TTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCAC 317
             |||  ||  ||||| ||||   ||||  |      |||| | |  | |||| |||  |
Db      1244 ATGTGAAGAATGTGGCAAAGCCTTTAACACATCCTCACACCTCACCACACATAAAAGAAT 1303

Qy       318 CCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACA 374
             |||||               | | || |||| | |||    |||| ||| | |
Db      1304 TCATACCGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTTCTC 1363

Qy       375 TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATG 434
             |||  |   | |||| | ||| |   |  |   || ||| | | |  |||||| ||||
Db      1364 ACAACTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAGAATG 1423

Qy       435 TAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTG 493
             |  |||| ||| |  ||  | |||||  |  |  ||||||   |||  |||  ||| |
Db      1424 TGGCAAAGCTTTTAAGCGGTCCTCAAACCTTACTGAACATAGGATAATTCATACTGGAG 1482
```

RESULT 10
AAS64586
ID AAS64586 standard; cDNA; 3639 BP.
XX
AC AAS64586;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #390.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG00399.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 390; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal

```
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 92.4;  DB 5;  Length 3639;
  Best Local Similarity   54.1%;  Pred. No. 1.2e-08;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||  |   | |||    | |
Db        738 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 797

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db        798 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 851

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              ||||| |||  | |    |||    | |  | |||||    |   ||| | | | | ||
Db        852 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 911

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  ||||| ||||   |||    |     | | |   | ||
Db        912 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 971

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |         | | || | || | ||| |  |||| ||| |
Db        972 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAGCTTTTAA 1031

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                ||    | |   |   | |||| | ||| ||  || |   || |||    |  |  ||
Db       1032 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1091

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |  |   |||||| || |  ||
Db       1092 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1151

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db       1152 TACTGGAGAGAAACCT 1167


RESULT 11
ACH87568/c
ID   ACH87568 standard; DNA; 994 BP.
XX
AC   ACH87568;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon probe #20763.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 1; SEQ ID NO 20763; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 994 BP; 252 A; 159 C; 203 G; 380 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 92;  DB 12;  Length 994;
  Best Local Similarity   52.0%;  Pred. No. 1.2e-08;
  Matches  259;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |     || ||     |  || |   |     |   | | |
Db        763 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 704

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |    | || || |||||  |||| ||||||  | ||| |  || |||    |
Db        703 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 644

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                  ||||      |  ||   ||| ||   ||| || ||||||      || || |||
Db        643 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 590
```

```
Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |     | |||||   ||   ||||| | | |
Db        589 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 530

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||  |   ||||| |||||  |||| |        |  | | | || |
Db        529 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 470

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| |||  |  |||||            |  | || |||| | |||    |||| |||
Db        469 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 410

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               ||     |  ||   |  | |||| | ||  |   |  |   || ||| | |  |
Db        409 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 350

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| || | |   | |||||| |  |   ||||||  |||
Db        349 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 290

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |    |||
Db        289 CATACTGGAGAGAAACCT 272
```

RESULT 12
ADR14368
ID ADR14368 standard; DNA; 2873 BP.
XX
AC ADR14368;
XX
DT 21-OCT-2004 (first entry)
XX
DE Human NF-kappaB pathway-associated gene SeqID369.
XX
KW NF-kappaB pathway; antiinflammatory; cytostatic; hepatotropic; virucide;
KW antiarthritic; antirheumatic; gastrointestinal-Gen; antiasthmatic;
KW antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;
KW immunosuppressive; vulnerary; gene therapy; immune disorder;
KW inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;
KW hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;
KW hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;
KW X-linked anhidrotic ectodermal dysplasia; immunodeficiency;
KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;
KW viral replication; host cell survival; evasion of immune response;
KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;
KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;
KW autoimmune disorder; hyper immune activity;
KW aberrant acute phase response; hypercongenital condition; birth defect;
KW necrotic lesion; wound; organ transplant rejection;
KW aberrant signal transduction; proliferating disorder; cancer;
KW HIV propagation; gene; ds; human.
XX
OS Homo sapiens.
XX
PN WO2004065577-A2.
XX
PD 05-AUG-2004.
XX
PF 13-JAN-2004; 2004WO-US000798.
XX
PR 14-JAN-2003; 2003US-0440068P.
PR 12-MAY-2003; 2003US-0469757P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Nadler SG, Neubauer MG, Feder JN, Carman J;
XX
DR WPI; 2004-562168/54.
DR P-PSDB; ADR14369.
XX
PT New isolated polynucleotides and polypeptides associated with NF-kappaB
PT pathway, useful for diagnosing, treating, or preventing disorders or
PT diseases associated with NF-kappaB pathway.
XX
PS Claim 1; SEQ ID NO 369; 237pp; English.
XX
CC This invention relates to the novel association of protein sequences (and
CC the genes which encode them) to the NF-kappaB pathway. The invention may
CC be useful for the production of compounds with an antiinflammatory,
CC cytostatic, hepatotropic, virucide, antiarthritic, antirheumatic,
CC gastrointestinal-Gen, antiasthmatic, antiarteriosclerotic,
CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or
CC vulnerary activity or for gene therapy. The proteins and nucleotides are
CC useful for diagnosing, preventing, treating, or ameliorating conditions
CC or diseases associated with the NF-kappaB pathway. The condition is an
CC immune disorder, an inflammatory disorder, an inflammatory disorder
CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,
CC hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM
CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic
CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,
CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell
CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory
CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick
CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper
CC immune activity, disorders related to aberrant acute phase responses,
CC hypercongenital conditions, birth defects, necrotic lesions, wounds,
CC organ transplant rejection, conditions related to organ transplant
CC rejection, disorders related to aberrant signal transduction,
CC proliferating disorders, cancers and HIV propagation in cells infected
CC with other viruses. The present sequence is that of a human gene which is
CC subject to the novel association with the NF-kappaB pathway of the
CC invention. Note: This sequence does not appear in the specification but
CC was obtained by the indexer from Genbank.
XX
SQ Sequence 2873 BP; 948 A; 564 C; 545 G; 816 T; 0 U; 0 Other;

Query Match 7.4%; Score 92; DB 13; Length 2873;
Best Local Similarity 52.2%; Pred. No. 1.4e-08;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||   || ||     |   |   |   |
Db       1621 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 1680

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | |  | || | | || || |||||  | ||  | ||||| |||  |  | |||    |
Db       1681 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 1740

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  |||||    ||||  || | |    ||   ||| || ||||||| |       || |||
Db       1741 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 1794

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |   |  |||||  |    | |||||    |   ||| | ||| |
Db       1795 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 1854

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  ||||| ||||   ||||       |      | | |  | |
Db       1855 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 1914

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||   |  | ||   |  | || ||| || |||     ||| |||
Db       1915 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1974

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
              |  ||      | | |  | ||||  |||  |  |       || |||   |
Db       1975 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 2034

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||    |||| ||| |  ||| | |||||| |  |  |||||||  |  |
Db       2035 TGTGAAGAATGCGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGTACATT 2094

Qy        484 CATCATGGTG 493
              |||  ||| |
```

```
Db       2095 CATACTGGAG 2104


RESULT 13
ADN99064
ID   ADN99064 standard; cDNA; 1533 BP.
XX
AC   ADN99064;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Novel human cDNA sequence #664.
XX
KW   ds; gene; anti-inflammatory; dermatological; neuroprotective;
KW   immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;
KW   gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW   psoriasis; diabetes; early aging; hormonal imbalance;
KW   ischemic heart disease; ulcerative colitis.
XX
OS   Homo sapiens.
XX
PN   WO2004038003-A2.
XX
PD   06-MAY-2004.
XX
PF   24-OCT-2003; 2003WO-US033947.
XX
PR   25-OCT-2002; 2002US-0421061P.
PR   25-OCT-2002; 2002US-0421080P.
PR   25-OCT-2002; 2002US-0421552P.
PR   25-OCT-2002; 2002US-0421614P.
PR   30-OCT-2002; 2002US-0422177P.
PR   30-OCT-2002; 2002US-0422178P.
PR   15-NOV-2002; 2002US-0426355P.
PR   15-NOV-2002; 2002US-0426384P.
PR   15-NOV-2002; 2002US-0426394P.
PR   15-NOV-2002; 2002US-0426430P.
PR   15-NOV-2002; 2002US-0426916P.
PR   27-NOV-2002; 2002US-0429224P.
PR   27-NOV-2002; 2002US-0429275P.
PR   27-NOV-2002; 2002US-0429302P.
PR   27-NOV-2002; 2002US-0429326P.
PR   27-NOV-2002; 2002US-0429651P.
PR   04-DEC-2002; 2002US-0430645P.
PR   04-DEC-2002; 2002US-0430651P.
PR   04-DEC-2002; 2002US-0430657P.
PR   04-DEC-2002; 2002US-0430663P.
PR   04-DEC-2002; 2002US-0430668P.
PR   04-DEC-2002; 2002US-0430684P.
PR   05-DEC-2002; 2002US-0430937P.
PR   05-DEC-2002; 2002US-0430965P.
PR   05-DEC-2002; 2002US-0431458P.
PR   12-DEC-2002; 2002US-0433251P.
PR   12-DEC-2002; 2002US-0433500P.
PR   13-DEC-2002; 2002US-0433316P.
PR   13-DEC-2002; 2002US-0433318P.
PR   23-DEC-2002; 2002US-0436236P.
PR   03-JAN-2003; 2003US-0437914P.
PR   17-JAN-2003; 2003US-0440820P.
PR   17-JAN-2003; 2003US-0440821P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476621P.
PR   09-JUN-2003; 2003US-0476632P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485217P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   08-JUL-2003; 2003US-0485359P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Kothakota S,  Lin H,  Linnemann T,  Pierce K,  Wang Y;
PI   Wong JGP,  Wu G,  Zhang H,  Zeng C;
XX
DR   WPI; 2004-365511/34.
DR   P-PSDB; ADN99848.
XX
PT   New nucleic acid molecules, useful in preparing a composition for
PT   treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT   disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT   ulcerative colitis.
XX
PS   Claim 1; SEQ ID NO 664; 532pp; English.
XX
CC   The invention relates to a nucleic acid molecule comprising a
CC   polynucleotide sequence or its complement that encodes a polypeptide. The
CC   nucleic acid is useful in preparing a composition for treating or
CC   preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC   cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC   heart disease or ulcerative colitis. This sequence corresponds to a
CC   nucleic acid of the invention.
XX
SQ   Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 91.8;  DB 12;  Length 1533;
  Best Local Similarity   52.1%;  Pred. No. 1.4e-08;
  Matches  257;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy         11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
              ||| | || |     |  ||     ||   | || ||      |     |   | | | | |
Db        305 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 364

Qy         71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
              |||||  || || || |||||  |||| |||||||| |||  |  | |||    |
Db        365 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 424

Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
              |||||   | ||  || | |   ||   ||| || ||||||      || || ||||  |
Db        425 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 478

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
              |||| |||  | |    ||| || |    | |||||   ||   ||| | | | | ||||
Db        479 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 538

Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | |||  ||  ||||| ||||    ||||          |  | | |  | ||||
Db        539 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 598

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  |||||| |           |  | || |||| | |||    | || ||| |
Db        599 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 658

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
```

```
Db        659 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 718

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487

Db        719 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 778

Qy        488 ATGGTGGATCTCC 500

Db        779 CTGGGGAAAAACC 791
```

RESULT 14
ADO00633
ID ADO00633 standard; cDNA; 1533 BP.
XX
AC ADO00633;
XX
DT 29-JUL-2004 (first entry)
XX
DE Novel human cDNA sequence #1448.
XX
KW ds; gene; anti-inflammatory; dermatological; neuroprotective;
KW immunomodulator; antibacterial; virucide; antipsoriatic; cytostatic;
KW gene therapy; vaccine; inflammatory; CNS; immune disorder; cancer;
KW psoriasis; diabetes; early aging; hormonal imbalance;
KW ischemic heart disease; ulcerative colitis.
XX
OS Homo sapiens.
XX
PN WO2004038003-A2.
XX
PD 06-MAY-2004.
XX
PF 24-OCT-2003; 2003WO-US033947.
XX
PR 25-OCT-2002; 2002US-0421061P.
PR 25-OCT-2002; 2002US-0421080P.
PR 25-OCT-2002; 2002US-0421552P.
PR 25-OCT-2002; 2002US-0421614P.
PR 30-OCT-2002; 2002US-0422177P.
PR 30-OCT-2002; 2002US-0422178P.
PR 15-NOV-2002; 2002US-0426355P.
PR 15-NOV-2002; 2002US-0426384P.
PR 15-NOV-2002; 2002US-0426394P.
PR 15-NOV-2002; 2002US-0426430P.
PR 15-NOV-2002; 2002US-0426916P.
PR 27-NOV-2002; 2002US-0429224P.
PR 27-NOV-2002; 2002US-0429275P.
PR 27-NOV-2002; 2002US-0429302P.
PR 27-NOV-2002; 2002US-0429326P.
PR 27-NOV-2002; 2002US-0429651P.
PR 04-DEC-2002; 2002US-0430645P.
PR 04-DEC-2002; 2002US-0430651P.
PR 04-DEC-2002; 2002US-0430657P.
PR 04-DEC-2002; 2002US-0430663P.
PR 04-DEC-2002; 2002US-0430668P.
PR 04-DEC-2002; 2002US-0430684P.
PR 05-DEC-2002; 2002US-0430937P.
PR 05-DEC-2002; 2002US-0430965P.
PR 05-DEC-2002; 2002US-0431458P.
PR 12-DEC-2002; 2002US-0433251P.
PR 12-DEC-2002; 2002US-0433500P.
PR 13-DEC-2002; 2002US-0433316P.
PR 13-DEC-2002; 2002US-0433318P.
PR 23-DEC-2002; 2002US-0436236P.
PR 03-JAN-2003; 2003US-0437914P.
PR 17-JAN-2003; 2003US-0440820P.
PR 17-JAN-2003; 2003US-0440821P.
PR 18-APR-2003; 2003US-0463700P.
PR 18-APR-2003; 2003US-0463708P.
PR 18-APR-2003; 2003US-0463716P.
PR 18-APR-2003; 2003US-0463732P.
PR 02-MAY-2003; 2003US-0467199P.
PR 02-MAY-2003; 2003US-0467201P.
PR 02-MAY-2003; 2003US-0467203P.
PR 02-MAY-2003; 2003US-0467230P.
PR 19-MAY-2003; 2003US-0471306P.
PR 19-MAY-2003; 2003US-0471336P.
PR 22-MAY-2003; 2003US-0472420P.
PR 22-MAY-2003; 2003US-0472430P.
PR 09-JUN-2003; 2003US-0476609P.
PR 09-JUN-2003; 2003US-0476621P.
PR 09-JUN-2003; 2003US-0476632P.
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485217P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485325P.
PR 08-JUL-2003; 2003US-0485359P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;
PI Wong JGP, Wu G, Zhang H, Zeng C;
XX
DR WPI; 2004-365511/34.
DR P-PSDB; ADN99848.
XX
PT New nucleic acid molecules, useful in preparing a composition for
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT ulcerative colitis.
XX
PS Claim 1; SEQ ID NO 2232; 532pp; English.
XX
CC The invention relates to a nucleic acid molecule comprising a
CC polynucleotide sequence or its complement that encodes a polypeptide. The
CC nucleic acid is useful in preparing a composition for treating or
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC heart disease or ulcerative colitis. This sequence corresponds to a
CC nucleic acid of the invention.
XX
SQ Sequence 1533 BP; 614 A; 280 C; 251 G; 388 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 12; Length 1533;
Best Local Similarity 52.1%; Pred. No. 1.4e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

```
Qy         11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70

Db        305 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 364

Qy         71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130

Db        365 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 424

Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190

Db        425 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 478

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250

Db        479 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 538
```

```
Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
Db        539 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 598

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
Db        599 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 658

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
Db        659 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 718

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
Db        719 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 778

Qy        488 ATGGTGGATCTCC 500
Db        779 CTGGGGAAAAACC 791
```

RESULT 15
ADB63624
ID ADB63624 standard; cDNA; 2110 BP.
XX
AC ADB63624;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone THYMU20071120.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   170. .1792
FT   /*tag= a
FT   /product= "Clone THYMU20071120 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB65594.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 10; Length 2110;
Best Local Similarity 51.7%; Pred. No. 1.5e-08;
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

```
Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
Db       1253 CATACCGAAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAAACACTCCTCA 1312

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
Db       1313 GCCCTTACTACACATAAGAGAATTCACACTGGAGAGAAACCCTACAAATGT------GAA 1366

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
Db       1367 GAATGTGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGAAACTTCATACT 1426

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
Db       1427 GGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTATCCAATCCTCAAAACTT 1486

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
Db       1487 ACTGAACATAAAAAAATTCATTCTGGAGAGATACCCTACAAGTGTGAAGAATGTGGCAAA 1546

Qy        361 GCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAAAACA 417
Db       1547 GCTTTTAAACACTCCTCATCCCTTACTACACATAAAAGAATTCATACTGGGGAGAAACCC 1606

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
Db       1607 TACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGATCCTCAAAACTTACTGAACATAAG 1666

Qy        478 TTAAAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
Db       1667 ATAATTCATACTGGAGAGAAACCCTATAAATGTGAGAGATGTGACAAAGCTTTTAACCAA 1726

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
Db       1727 TCTGCAAACCT------TACTAAACATAAAAAAATACATACTGGAGAGAAACTACAGAAC 1780

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
Db       1781 TGGAATGTGTGATAATGATTTTGACAACACCTCAAATTTTTCTAAATATAAAGGAAATCA 1840

Qy        657 TGATGAT 663
Db       1841 TACTGGT 1847
```

Search completed: February 9, 2005, 02:39:51
Job time : 644.278 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:51:00 ; Search time 224.272 Seconds
(without alignments)
9039.698 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	584.8	47.2	627	4	US-09-248-796A-5184	Sequence 5184, Ap
	2	391	31.6	462	4	US-09-248-796A-5186	Sequence 5186, Ap
	3	96.6	7.8	2320	4	US-09-016-434-1054	Sequence 1054, Ap
	4	90.8	7.3	3798	4	US-09-949-016-4204	Sequence 4204, Ap
	5	90.8	7.3	156942	4	US-09-949-016-12227	Sequence 12227, A
	6	90.8	7.3	156950	4	US-09-949-016-15946	Sequence 15946, A
	7	89.2	7.2	3839	4	US-09-949-016-485	Sequence 485, App
c	8	86.8	7.0	601	4	US-09-949-016-27359	Sequence 27359, A
c	9	86.8	7.0	601	4	US-09-949-016-149638	Sequence 149638,
	10	86.2	7.0	601	4	US-09-949-016-44799	Sequence 44799, A
	11	86.2	7.0	45138	4	US-09-949-016-13027	Sequence 13027, A
c	12	85.6	6.9	601	4	US-09-949-016-27360	Sequence 27360, A
c	13	85.6	6.9	601	4	US-09-949-016-149637	Sequence 149637,
c	14	83.2	6.7	601	4	US-09-949-016-27361	Sequence 27361, A
c	15	83.2	6.7	601	4	US-09-949-016-149636	Sequence 149636,
	16	81.6	6.6	472	4	US-09-513-999C-12522	Sequence 12522, A
	17	80.2	6.5	2804	4	US-09-949-016-2278	Sequence 2278, Ap
	18	80.2	6.5	22294	4	US-09-949-016-14020	Sequence 14020, A
	19	79.4	6.4	1030	4	US-09-976-594-370	Sequence 370, App
	20	79.2	6.4	2982	4	US-09-949-016-4184	Sequence 4184, Ap
	21	79.2	6.4	41435	4	US-09-949-016-15926	Sequence 15926, A
	22	78.2	6.3	3400	4	US-09-799-451-136	Sequence 136, App
c	23	77.2	6.2	601	4	US-09-949-016-149473	Sequence 149473,
c	24	76	6.1	601	4	US-09-949-016-149474	Sequence 149474,
	25	76	6.1	2501	4	US-09-949-016-5363	Sequence 5363, Ap
	26	76	6.1	2504	4	US-09-949-016-487	Sequence 487, App
	27	76	6.1	29671	4	US-09-949-016-12229	Sequence 12229, A
	28	76	6.1	29671	4	US-09-949-016-17105	Sequence 17105, A
	29	74.4	6.0	2143	4	US-09-881-578A-3	Sequence 3, Appli
	30	73	5.9	2925	4	US-09-620-312D-163	Sequence 163, App
	31	71.2	5.7	979	4	US-09-976-594-939	Sequence 939, App
	32	69.6	5.6	2402	4	US-09-949-016-1710	Sequence 1710, Ap
	33	69.6	5.6	2407	4	US-09-949-016-489	Sequence 489, App
	34	69.6	5.6	30867	4	US-09-949-016-12231	Sequence 12231, A
	35	69.6	5.6	30868	4	US-09-949-016-13452	Sequence 13452, A
	36	68	5.5	2891	4	US-09-949-016-1115	Sequence 1115, Ap
c	37	67.6	5.5	601	4	US-09-949-016-39838	Sequence 39838, A
c	38	67.6	5.5	601	4	US-09-949-016-160861	Sequence 160861,
c	39	66.8	5.4	18773	4	US-09-949-016-14164	Sequence 14164, A
	40	66.4	5.4	2891	4	US-09-949-016-4503	Sequence 4503, Ap
	41	66.4	5.4	21360	4	US-09-949-016-12857	Sequence 12857, A
	42	66.4	5.4	21361	4	US-09-949-016-16245	Sequence 16245, A
	43	66.2	5.3	2752	4	US-09-949-016-3563	Sequence 3563, Ap
	44	66.2	5.3	18824	4	US-09-949-016-15305	Sequence 15305, A
	45	65.4	5.3	2274	4	US-09-620-312D-154	Sequence 154, App

ALIGNMENTS

RESULT 1
US-09-248-796A-5184
; Sequence 5184, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5184
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5184

Query Match 47.2%; Score 584.8; DB 4; Length 627;
Best Local Similarity 97.4%; Pred. No. 2.5e-124;
Matches 608; Conservative 0; Mismatches 7; Indels 9; Gaps 1;

```
Qy        616 TGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATC 675
              ||| | || |||||||||||||||||||||||||||||||||||||||||| |||||||||
Db         13 TGTTTGGGAAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTTTACCATGATC 72

Qy        676 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 AAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTT 132

Qy        736 GAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTATTAAAGGAAACTGAA 795
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        133 GAACATTATAATATCTTCCATGATGGTAATATCCCCGATGATTTATTAAAGGAAACTGAA 192

Qy        796 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAATAATTTGCATGAATTA 252

Qy        856 GAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAAGATAGTCTAGATGAA 915
              |||||||||||||||||||          ||||||||||||||||||||||||||||||
Db        253 GAAACAGAGAAATTAAAAGTG---------GAAGAAGATGAAGAAGATAGTCTAGATGAA 303

Qy        916 AAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 975
              | |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        304 AGAAGAAGTGATGTTAGATCAGATTCAATGTCAGCTCAAAGATCAATAAAATCATTTACT 363
```

```
Qy        976 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAATAGTGGGAAGAAGATC 423

Qy       1036 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATATGATTTACGTCGACAT 483

Qy       1096 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 TTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTAAATAGTATAGAAAAA 543

Qy       1156 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 1215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATGGATTTATTGCCAAAT 603

Qy       1216 GAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||
Db        604 GAAACATCAGTGATTTCTCGATAA 627
```

RESULT 2
US-09-248-796A-5186
; Sequence 5186, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5186
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5186

Query Match 31.6%; Score 391; DB 4; Length 462;
Best Local Similarity 98.5%; Pred. No. 4.4e-80;
Matches 405; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         52 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 111

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        112 CGTCCCAAAAAGTATATTTGTACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 171

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 231

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 291

Qy        241 GAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 GAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 351

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        352 AAAAGACATGAAATCACCCATACCAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 411

Qy        361 GCATTTTAT-AAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410
              |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        412 ACATTTTATAAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 462
```

RESULT 3
US-09-016-434-1054
; Sequence 1054, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1054:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1017721
US-09-016-434-1054

Query Match 7.8%; Score 96.6; DB 4; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.4e-12;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || | |||  |||| |||||| | ||| ||| | |||    |
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                  |||  || ||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               ||||| |||          ||| |  |    |||||||   ||   ||| | | | | ||
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  |   ||||| ||||   ||||          |||| | |  | |||
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              |||  | |  |||||             |  | || |||||| |||  | |||| ||| |
```

```
Db       1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy        369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | |     ||   |   | |||| | ||| |   |  |    || ||| | |   |  ||
Db       1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              ||||  |||||  ||||| ||| |   |    ||||||    |   |  |||||||  |||  ||
Db       1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy        486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
              |  ||| |      |||     |         ||    |         ||||||  | | |  ||
Db       1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy        546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
              | | |    |      |    |     | | | ||  |   | |||||   |||||
Db       1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy        606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||| |||| || |      |     | | |  ||  ||||| |  |||||||   ||
Db       1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551


RESULT 4
US-09-949-016-4204
; Sequence 4204, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4204
;   LENGTH: 3798
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-4204

  Query Match             7.3%;  Score 90.8;  DB 4;  Length 3798;
  Best Local Similarity   53.9%;  Pred. No. 3.4e-11;
  Matches  235;  Conservative    0;  Mismatches  192;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  |||| | ||   |    | |||    | |
Db       1848 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1907

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||       || || ||||
Db       1908 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1961

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  | |     |||     | |  | ||||      |   ||| | | | ||
Db       1962 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 2021

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | | | |||  ||  ||||| ||||    |||     | | |    |  ||
Db       2022 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 2081

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |           |  | ||| | || | |||  |  || | ||| |
Db       2082 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 2141

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              ||     | |   |   | |||| | ||| ||  || |   || |||   |  |  ||
Db       2142 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 2201

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |   |   ||||| || |  ||
Db       2202 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 2261

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db       2262 TACTGGAGAGAAACCT 2277


RESULT 5
US-09-949-016-12227
; Sequence 12227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12227
;   LENGTH: 156942
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(156942)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12227

  Query Match             7.3%;  Score 90.8;  DB 4;  Length 156942;
  Best Local Similarity   53.9%;  Pred. No. 1.1e-10;
  Matches  235;  Conservative    0;  Mismatches  192;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||   |    | |||    | |
Db     152999 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 153058

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||       || || ||||
Db     153059 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 153112

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  | |     |||     | |  | ||||      |   ||| | | | ||
Db     153113 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 153172

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | | | |||  ||  ||||| ||||    |||     | | |    |  ||
Db     153173 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 153232

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |           |  | ||| | || | |||  |  || | ||| |
Db     153233 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 153292

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              ||     | |   |   | |||| | ||| ||  || |   || |||   |  |  ||
Db     153293 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 153352

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |   |   ||||| || |  ||
Db     153353 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 153412
```

```
Qy        486 TCATGGTGGATCTCCT 501
              |  |||  |     |||
Db     153413 TACTGGAGAGAAACCT 153428


RESULT 6
US-09-949-016-15946
; Sequence 15946, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15946
;   LENGTH: 156950
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(156950)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15946

  Query Match             7.3%;  Score 90.8;  DB 4;  Length 156950;
  Best Local Similarity   53.9%;  Pred. No. 1.1e-10;
  Matches  235;  Conservative    0;  Mismatches  192;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||  |    | |||     | |
Db     152999 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 153058

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db     153059 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 153112

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  | |    |||    | |  | |||||    |   ||| | | | | ||
Db     153113 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 153172

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | |  |||   ||  |||||  ||||   |||    |      | | |    | ||
Db     153173 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 153232

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | ||  ||  |||||  |            |  | ||  | ||  | |||  |  || | ||| |
Db     153233 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 153292

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              ||    | |   |   | |||| | |||  ||  ||  |   ||  |||   |  |  ||
Db     153293 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 153352

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | ||  |||||   |||| ||| | |||| |||||||  |  |   ||||||  || |  ||
Db     153353 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 153412

Qy        486 TCATGGTGGATCTCCT 501
              |  |||  |     |||
Db     153413 TACTGGAGAGAAACCT 153428


RESULT 7
US-09-949-016-485
; Sequence 485, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 485
;   LENGTH: 3839
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-485

  Query Match             7.2%;  Score 89.2;  DB 4;  Length 3839;
  Best Local Similarity   53.7%;  Pred. No. 8e-11;
  Matches  234;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||  |    | |||     | |
Db        964 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  | |    |||    | |  | |||||    |   ||| | | | | ||
Db       1078 CAAAGCTTTTAGCCATTCTTCAGCCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  | |  |||   ||  |||||  ||||   |||    |      | | |    | ||
Db       1138 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | ||  ||  |||||  |            |  | ||  | ||  | |||  |  || | ||| |
Db       1198 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 1257

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              ||    | |   |   | |||| | |||  ||  ||  |   ||  |||   |  |  ||
Db       1258 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | ||  |||||   |||| ||| | |||| |||||||  |  |   ||||||  || |  ||
Db       1318 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1377

Qy        486 TCATGGTGGATCTCCT 501
              |  |||  |     |||
Db       1378 TACTGGAGAGAAACCT 1393


RESULT 8
US-09-949-016-27359/c
; Sequence 27359, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
```

```
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27359
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-27359

  Query Match             7.0%;  Score 86.8;  DB 4;  Length 601;
  Best Local Similarity   51.4%;  Pred. No. 1.6e-10;
  Matches  252;  Conservative    1;  Mismatches  228;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |       |   || ||      |  ||      |      |  || | |
Db        538 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 479

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||    |
Db        478 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 419

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |  ||||    | ||  || | ||| ||   ||| |  ||  ||      || ||| |||
Db        418 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 365

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||||||  || |    |||     |      |||||    |   ||| ||  | |
Db        364 GACAAAGCATTTATCTGGTCCTCAACCCTTACTGAACATAAGAGAATTCATACCAGAGAG 305

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||:| | | | |||  ||  ||||| ||||   |||      |    ||||| | | | |
Db        304 AAAMCCTACAAATGTGAAGAATGTGGCAAAGCATTTAGCCAGCCTTCACACCTTACTACA 245

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |   | || ||                | | || |||| | |||    |||| |||
Db        244 CATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTT 185

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                ||      | |   |   | |||| | ||| |   |  |   || ||| | |  |
Db        184 AGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAA 125

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| ||| |    | ||||||   |  |  ||||||| |||
Db        124 TGTGAAGAATGTGGCAAAGCTTTTAGGAAATCTTCAACTCTTACTGAACATAAGATAATT 65

Qy        484 CATCATGGTG 493
              |||  ||| |
Db         64 CATACTGGAG 55


RESULT 9
US-09-949-016-149638/c
; Sequence 149638, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149638
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-149638

  Query Match             7.0%;  Score 86.8;  DB 4;  Length 601;
  Best Local Similarity   51.4%;  Pred. No. 1.6e-10;
  Matches  252;  Conservative    1;  Mismatches  228;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |       |   || ||      |  ||      |      |  || | |
Db        538 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 479

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||    |
Db        478 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 419

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |  ||||    | ||  || | ||| ||   ||| |  ||  ||      || ||| |||
Db        418 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 365

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||||||  || |    |||     |      |||||    |   ||| ||  | |
Db        364 GACAAAGCATTTATCTGGTCCTCAACCCTTACTGAACATAAGAGAATTCATACCAGAGAG 305

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||:| | | | |||  ||  ||||| ||||   |||      |    ||||| | | | |
Db        304 AAAMCCTACAAATGTGAAGAATGTGGCAAAGCATTTAGCCAGCCTTCACACCTTACTACA 245

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |   | || ||                | | || |||| | |||    |||| |||
Db        244 CATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTT 185

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                ||      | |   |   | |||| | ||| |   |  |   || ||| | |  |
Db        184 AGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAA 125

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| ||| |    | ||||||   |  |  ||||||| |||
Db        124 TGTGAAGAATGTGGCAAAGCTTTTAGGAAATCTTCAACTCTTACTGAACATAAGATAATT 65

Qy        484 CATCATGGTG 493
              |||  ||| |
Db         64 CATACTGGAG 55


RESULT 10
US-09-949-016-44799
; Sequence 44799, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44799
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-44799

  Query Match             7.0%;  Score 86.2;  DB 4;  Length 601;
  Best Local Similarity   54.4%;  Pred. No. 2.2e-10;
```

```
  Matches  241;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    3;

Qy          61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
               | | | | | || |    | || || |||||  |||| || || | ||   | | |||
Db          56 CATACTAGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCATTTATATGGTCCTCA 115

Qy         121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                 ||     ||||   || |  |||| |   ||   ||| || ||||||      ||
Db         116 ACCCTAACTAGACATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGT------GAA 169

Qy         181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
               || ||||  ||| | |||  |  |   |||     |    |||||||   ||   ||| |
Db         170 GAATGTGGCAAATCTTTTAGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACT 229

Qy         241 GAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
               | | | ||||| | | | |||  ||  ||||| ||||   ||||  |       |    |
Db         230 GGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACTGGTCCTCAACTCTT 289

Qy         301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
               |  | |||| | || |  |||||  |            | | || ||||||  |||    ||
Db         290 ACTAAACATAASATAATTCATACTGAAGAGAAACCCTACAAATGTGAAAAATGTGGCAAA 349

Qy         361 GCATTTTA-TAAACATCAATCTTTA--AGACATCATATATTATCTGTTCATGAAAAAACA 417
               || ||| |  |  | ||||||  |||  |  ||| | | | |   |  |   || ||| |
Db         350 GCCTTTAAGCAGTCTTCAATCCTTACTAACCATAAGAGAATTCATACTGGAGAGAAACCC 409

Qy         418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
               |  |  ||| || |||||   |||  ||| |  ||  ||||||  ||  |  |||||||
Db         410 TATAAATGTGAAGAATGTGGCAAATCTTTTAACCGGTCTTCAACTTTTACTAAACATAAG 469

Qy         478 TTAAAACATCATGGTGGATCTCC 500
                 |||  |||  ||| | |   ||
Db         470 GTAATTCATACTGGAGTAAAACC 492


RESULT 11
US-09-949-016-13027
; Sequence 13027, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13027
;   LENGTH: 45138
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(45138)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13027

  Query Match             7.0%;  Score 86.2;  DB 4;  Length 45138;
  Best Local Similarity   54.4%;  Pred. No. 8.4e-10;
  Matches  241;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    3;

Qy          61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
               | | | | | || |    | || || |||||  |||| || || | ||   | | |||
Db       42121 CATACTAGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCATTTATATGGTCCTCA 42180

Qy         121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                 ||     ||||   || |  |||| |   ||   ||| || ||||||      ||
Db       42181 ACCCTAACTAGACATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGT------GAA 42234

Qy         181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
               || ||||  ||| | |||  |  |   |||     |    |||||||   ||   ||| |
Db       42235 GAATGTGGCAAATCTTTTAGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACT 42294

Qy         241 GAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
               | | | ||||| | | | |||  ||  ||||| ||||   ||||  |       |    |
Db       42295 GGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACTGGTCCTCAACTCTT 42354

Qy         301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
               |  | |||| | || |  |||||  |            | | || ||||||  |||    ||
Db       42355 ACTAAACATAAGATAATTCATACTGAAGAGAAACCCTACAAATGTGAAAAATGTGGCAAA 42414

Qy         361 GCATTTTA-TAAACATCAATCTTTA--AGACATCATATATTATCTGTTCATGAAAAAACA 417
               || ||| |  |  | ||||||  |||  |  ||| | | | |   |  |   || ||| |
Db       42415 GCCTTTAAGCAGTCTTCAATCCTTACTAACCATAAGAGAATTCATACTGGAGAGAAACCC 42474

Qy         418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
               |  |  ||| || |||||   |||  ||| |  ||  ||||||  ||  |  |||||||
Db       42475 TATAAATGTGAAGAATGTGGCAAATCTTTTAACCGGTCTTCAACTTTTACTAAACATAAG 42534

Qy         478 TTAAAACATCATGGTGGATCTCC 500
                 |||  |||  ||| | |   ||
Db       42535 GTAATTCATACTGGAGTAAAACC 42557


RESULT 12
US-09-949-016-27360/c
; Sequence 27360, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27360
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-27360

  Query Match             6.9%;  Score 85.6;  DB 4;  Length 601;
  Best Local Similarity   51.4%;  Pred. No. 3e-10;
  Matches  252;  Conservative    0;  Mismatches  229;  Indels    9;  Gaps    2;

Qy           7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
               ||| ||| | || |      |   || ||     |  ||     |    |  || | |
Db         499 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 440

Qy          67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||    |
Db         439 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 380

Qy         127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |  ||||   | ||  || | ||| ||   ||| |  ||  ||      || || |||
Db         379 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 326

Qy         187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
               || ||||||||| || |    |||   |    |||||    |   ||| ||  | |
Db         325 GACAAAGCATTTATCTGGTCCTCARCCCTTACTGAACATAAGAGAATTCATACCAGAGAG 266
```

```
Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | ||| || ||||| ||||   |||      |    |||| | | | | |
Db        265 AAAACCTACAAATGTGAAGAATGTGGCAAAGCATTTAGCCAGCCTTCACACCTTACTACA 206

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |   | || ||                 | | || |||| | |||    |||| |||
Db        205 CATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTT 146

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                  ||      | |   |   | |||| | ||| |   |  |   || ||| | |  |
Db        145 AGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAA 86

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||||   |||| ||| |    | ||||||   |  |  |||||||  |||
Db         85 TGTGAAGAATGTGGCAAAGCTTTTAGGAAATCTTCAACTCTTACTGAACATAAGATAATT 26

Qy        484 CATCATGGTG 493
              |||  ||| |
Db         25 CATACTGGAG 16
```

RESULT 13
US-09-949-016-149637/c
; Sequence 149637, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149637
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149637

Query Match 6.9%; Score 85.6; DB 4; Length 601;
Best Local Similarity 51.4%; Pred. No. 3e-10;
Matches 252; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |       |   || ||      |  ||      |      |  || | |
Db        499 GAATGTGGCAAAGCTTTTAAGCACTCCTCAGCCCTTGCTAAACATAAAATAATACATGCT 440

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |  | | || || ||| |  ||||  ||||| | ||||| | | |||     |
Db        439 GGAGAGAAACTCTACAAATGTGAGGAATGTGGCAAAGCTTTTAATCAATCTTCAAATCTT 380

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  ||||   | ||  || | ||| ||   ||| |  ||  ||      || || |||
Db        379 ACGACACATAAGATAATTCATACTAAAGAGAAACCTTCCAAGAGT------GAAGAATGT 326

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||||||| || |     |||     |      |||||     |   ||| ||  | |
Db        325 GACAAAGCATTTATCTGGTCCTCARCCCTTACTGAACATAAGAGAATTCATACCAGAGAG 266

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | ||| || ||||| ||||   |||      |    |||| | | | | |
Db        265 AAAACCTACAAATGTGAAGAATGTGGCAAAGCATTTAGCCAGCCTTCACACCTTACTACA 206

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |   | || ||                 |  | || |||| | |||    |||| |||
Db        205 CATAAGAGGATGCACACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTT 146
```

```
Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                  ||      | |   |   | |||| | ||| |   |  |   || ||| | |  |
Db        145 AGCCAATCCTCAACCCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAA 86

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||||   |||| ||| |    | |||||||  |  |  |||||||  |||
Db         85 TGTGAAGAATGTGGCAAAGCTTTTAGGAAATCTTCAACTCTTACTGAACATAAGATAATT 26

Qy        484 CATCATGGTG 493
              |||  ||| |
Db         25 CATACTGGAG 16
```

RESULT 14
US-09-949-016-27361/c
; Sequence 27361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27361
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-27361

Query Match 6.7%; Score 83.2; DB 4; Length 601;
Best Local Similarity 52.5%; Pred. No. 1e-09;
Matches 231; Conservative 1; Mismatches 199; Indels 9; Gaps 2;

```
Qy         57 ATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACC 116
              | | | | |  || || |     | || || | |||  ||||| ||| | | ||| ||||
Db        554 AACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAAGCGACT 495

Qy        117 ATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGT 176
              |||    |      |||||  || ||  ||    |   ||   ||  || ||||||
Db        494 CTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATGT----- 440

Qy        177 GGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACA 236
              || || ||||  ||||| |||     ||   ||| || |    | |||||    |     ||
Db        439 -GAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 381

Qy        237 TTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACA 296
              | | | | | ||||| | | | |||  ||  ||||| ||||   |||| |        |
Db        380 TACTGGAGAGAAACCTTACAAGTGTGAAGAATGTGGCAAAGCATTTAACTGGTCCTCAAG 321

Qy        297 CTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCA 356
              | | |  | |||| |||  :  ||||| |              | | || | || | |||
Db        320 CCTTACTAAACATAAAAGAWTTCATACTAGAGAGAAACCCTTCAAATGTAAAGAATGTGG 261

Qy        357 AGAAGCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAA 413
              ||||||||          |  | |  ||  ||||||  | |    ||     |   || ||
Db        260 CAAAGCATTTATATGGTCTTCAACCCTAACTAGACATAAGAGGATACACACTGGAGAGAA 201

Qy        414 AACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACA 473
              | |  |  ||| || |||||   |||| ||| |  ||  | ||||    |  |  | ||
Db        200 GCCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGTTCCTCAACCCTTACTAAGCA 141

Qy        474 TAAATTAAAACATCATGGTG 493
              |||   ||  |||  ||| |
```

```
Db        140 TAAGACAATTCATACTGGAG 121


RESULT 15
US-09-949-016-149636/c
; Sequence 149636, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149636
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-149636

  Query Match             6.7%;  Score 83.2;  DB 4;  Length 601;
  Best Local Similarity   52.5%;  Pred. No. 1e-09;
  Matches  231;  Conservative    1;  Mismatches  199;  Indels    9;  Gaps    2;

Qy         57 ATCACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACC 116
              | | | | |  || || |    | || || | |||  ||||| ||| | | ||| ||||
Db        554 AACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAAGCGACT 495

Qy        117 ATCATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGT 176
               |||    |     |||||  || ||  ||   |   ||   ||  || ||||||
Db        494 CTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATGT----- 440

Qy        177 GGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACA 236
               || || ||||  ||||| |||    ||   ||| || |    | |||||   |    ||
Db        439 -GAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 381

Qy        237 TTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACA 296
              | | | | |||||  | | | |||  ||  ||||| ||||   |||| |       |
Db        380 TACTGGAGAGAAACCTTACAAGTGTGAAGAATGTGGCAAAGCATTTAACTGGTCCTCAAG 321

Qy        297 CTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCA 356
              | | |  | |||| |||  :  ||||| |           |  | || | || | |||
Db        320 CCTTACTAAACATAAAAGAWTTCATACTAGAGAGAAACCCTTCAAATGTAAAGAATGTGG 261

Qy        357 AGAAGCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAA 413
               ||||||||        |  | |  ||   |||||| | |   ||     |   || ||
Db        260 CAAAGCATTTATATGGTCTTCAACCCTAACTAGACATAAGAGGATACACACTGGAGAGAA 201

Qy        414 AACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACA 473
               | |  |  ||| || |||||   |||| ||| |  ||  | ||||   |  |  | ||
Db        200 GCCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGTTCCTCAACCCTTACTAAGCA 141

Qy        474 TAAATTAAAACATCATGGTG 493
              |||   ||  |||  ||| |
Db        140 TAAGACAATTCATACTGGAG 121


Search completed: February  9, 2005, 07:51:11
Job time : 226.272 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 05:36:40 ; Search time 2160.11 Seconds
(without alignments)
3301.382 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4313806 seqs, 2877871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1235.8	99.7	1239	16	US-10-032-585-6493	Sequence 6493, Ap
2	99.4	8.0	2132	17	US-10-094-749-726	Sequence 726, App
3	96.6	7.8	2320	17	US-10-305-720-1054	Sequence 1054, Ap
c 4	92	7.4	994	16	US-10-029-386-20763	Sequence 20763, A
5	92	7.4	2873	18	US-10-755-889-369	Sequence 369, App
6	91.8	7.4	2110	17	US-10-104-047-1778	Sequence 1778, Ap
7	91.8	7.4	3078	17	US-10-104-047-622	Sequence 622, App
8	90.4	7.3	2114	17	US-10-094-749-692	Sequence 692, App
9	90.2	7.3	2230	17	US-10-094-749-499	Sequence 499, App
10	89.8	7.2	590	16	US-10-029-386-13298	Sequence 13298, A
11	89.2	7.2	2221	17	US-10-108-260A-29	Sequence 29, Appl
12	89.2	7.2	3839	17	US-10-172-118-910	Sequence 910, App
13	89.2	7.2	3839	17	US-10-342-887-910	Sequence 910, App
14	89	7.2	3502	13	US-10-098-841-48	Sequence 48, Appl
15	88.6	7.2	1466	16	US-10-029-386-25074	Sequence 25074, A
c 16	88.4	7.1	1410	16	US-10-029-386-20532	Sequence 20532, A
17	88.4	7.1	2509	17	US-10-381-327-11	Sequence 11, Appl
18	88.4	7.1	4132	14	US-10-198-846-13133	Sequence 13133, A
19	87.6	7.1	583	16	US-10-029-386-13093	Sequence 13093, A
20	87.6	7.1	1757	18	US-10-203-052B-6	Sequence 6, Appli
21	87.2	7.0	831	17	US-10-264-049-975	Sequence 975, App
22	86.8	7.0	2662	17	US-10-363-616-66	Sequence 66, Appl
23	86.2	7.0	911	16	US-10-029-386-22636	Sequence 22636, A
24	86.2	7.0	1422	16	US-10-029-386-20322	Sequence 20322, A
25	86.2	7.0	1725	16	US-10-029-386-20556	Sequence 20556, A
c 26	86.2	7.0	1727	16	US-10-029-386-20367	Sequence 20367, A
27	86.2	7.0	2558	9	US-09-764-864-1664	Sequence 1664, Ap
c 28	86.2	7.0	2558	9	US-09-764-864-1665	Sequence 1665, Ap
29	86	6.9	2664	17	US-10-108-260A-511	Sequence 511, App
30	86	6.9	2690	18	US-10-723-860-8039	Sequence 8039, Ap
31	86	6.9	3400	17	US-10-104-047-1037	Sequence 1037, Ap
32	85.4	6.9	2505	17	US-10-108-260A-808	Sequence 808, App
33	84.8	6.8	2239	17	US-10-094-749-1084	Sequence 1084, Ap
34	84.4	6.8	1191	16	US-10-029-386-22992	Sequence 22992, A
35	84.4	6.8	3138	17	US-10-104-047-83	Sequence 83, Appl
c 36	84.2	6.8	1578	16	US-10-029-386-20214	Sequence 20214, A
37	84	6.8	565	16	US-10-029-386-3970	Sequence 3970, Ap
38	84	6.8	1619	18	US-10-422-522-46	Sequence 46, Appl
39	84	6.8	3617	16	US-10-029-386-22868	Sequence 22868, A
40	84	6.8	5820	17	US-10-363-616-44	Sequence 44, Appl
c 41	83.6	6.7	596	16	US-10-029-386-6664	Sequence 6664, Ap
c 42	83.2	6.7	1101	16	US-10-029-386-25405	Sequence 25405, A
43	83.2	6.7	2241	17	US-10-108-260A-1289	Sequence 1289, Ap
44	83.2	6.7	2525	17	US-10-108-260A-1358	Sequence 1358, Ap
c 45	83	6.7	500	16	US-10-029-386-6818	Sequence 6818, Ap

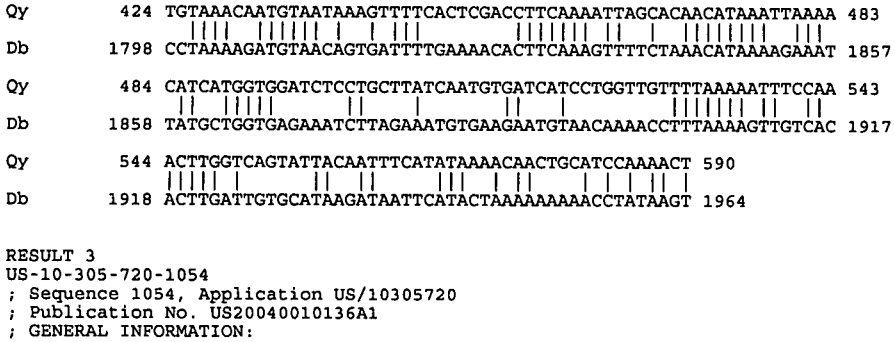
ALIGNMENTS

RESULT 1
US-10-032-585-6493
; Sequence 6493, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6493
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6493

Query Match 99.7%; Score 1235.8; DB 16; Length 1239;
Best Local Similarity 99.8%; Pred. No. 2.6e-227;
Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

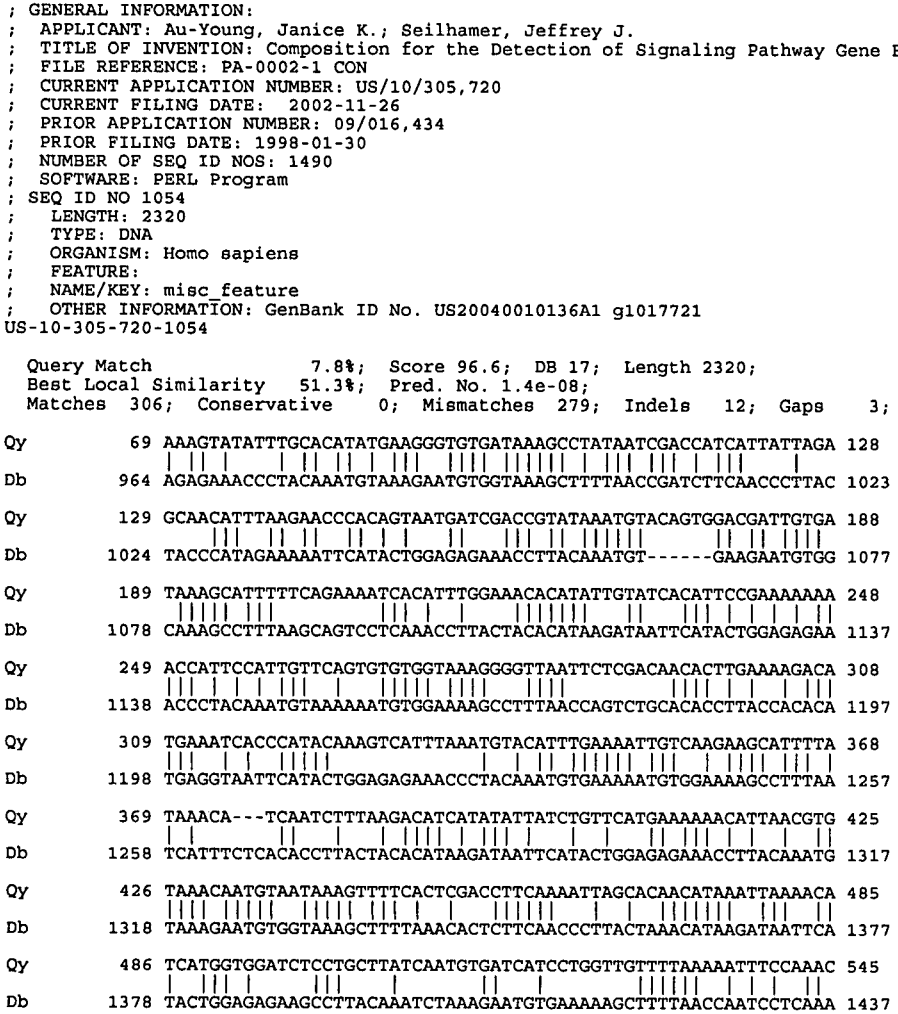
```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
```

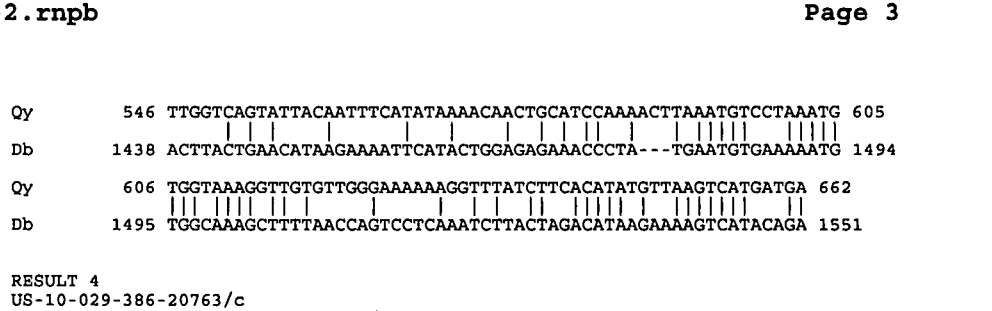
```
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239


RESULT 2
US-10-094-749-726
; Sequence 726, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 726
;   LENGTH: 2132
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-726

  Query Match             8.0%;  Score 99.4;  DB 17;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 4e-09;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |           || ||  |  |  || |   |      |   | | |
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |   | || || |||||  ||||  |||||  |   |||||  | |||  | |
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||    | ||  || | |    ||    ||| || | ||||      || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |       |||||  | |  |||||||  |||   ||| | | | |
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  ||||||||||    ||||       |||| | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  | ||||              |    || |||||| |||    |||| |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              | |     |   | |   ||||| | | | |   |  |  ||| |||   | |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797
```



RESULT 3
US-10-305-720-1054
; Sequence 1054, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expressic
; FILE REFERENCE: PA-0002-1 CON
; CURRENT APPLICATION NUMBER: US/10/305,720
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/016,434
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 1490
; SOFTWARE: PERL Program
; SEQ ID NO 1054
; LENGTH: 2320
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GenBank ID No. US20040010136A1 g1017721
US-10-305-720-1054

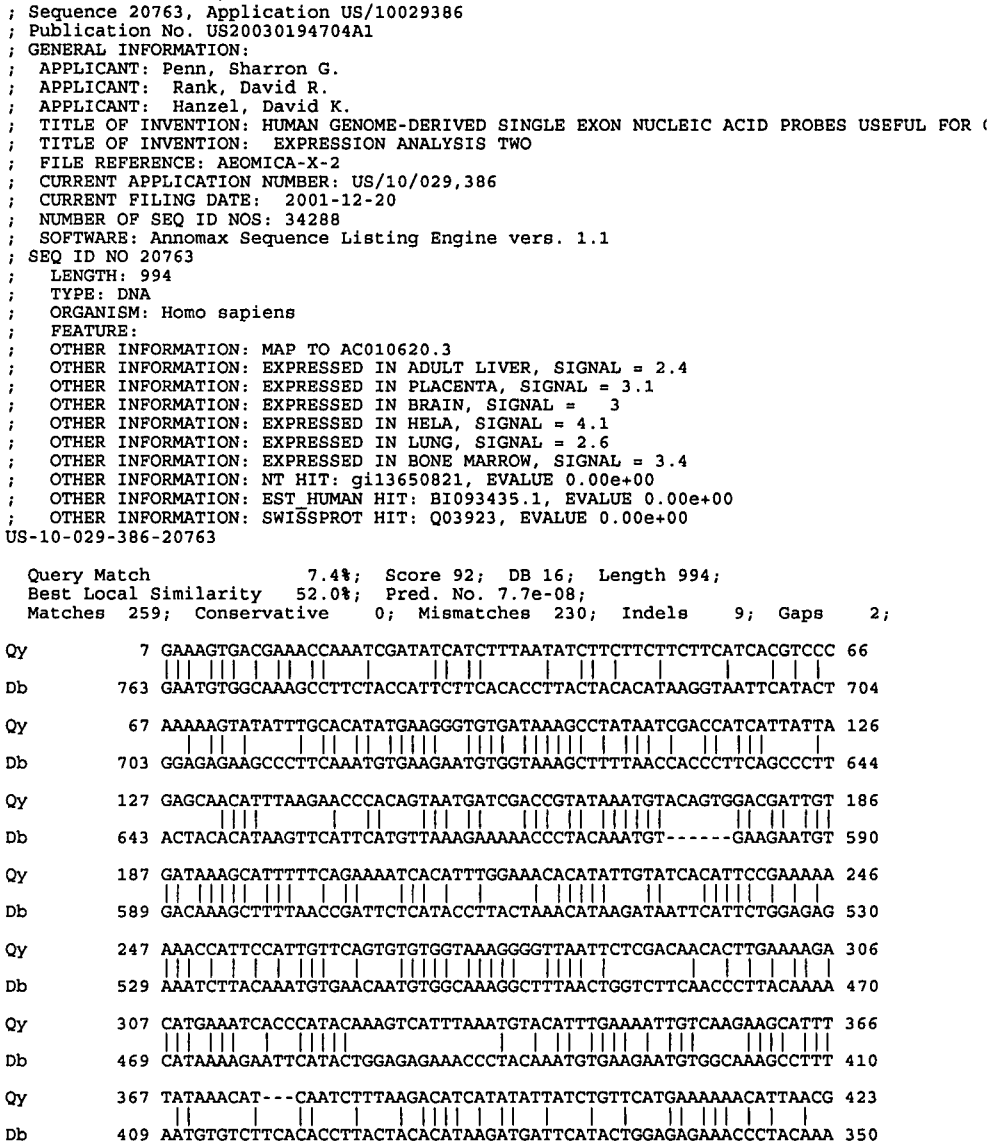
Query Match 7.8%; Score 96.6; DB 17; Length 2320;
Best Local Similarity 51.3%; Pred. No. 1.4e-08;
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;





RESULT 4
US-10-029-386-20763/c
; Sequence 20763, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR (
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20763
; LENGTH: 994
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010620.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: NT HIT: gi13650821, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BI093435.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUE 0.00e+00
US-10-029-386-20763

Query Match 7.4%; Score 92; DB 16; Length 994;
Best Local Similarity 52.0%; Pred. No. 7.7e-08;
Matches 259; Conservative 0; Mismatches 230; Indels 9; Gaps 2;



```
Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |  |   | ||||||  |  |   ||||||  |||
Db        349 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 290

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |     |||
Db        289 CATACTGGAGAGAAACCT 272
```

RESULT 5
US-10-755-889-369
; Sequence 369, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 369
; LENGTH: 2873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-369

Query Match 7.4%; Score 92; DB 18; Length 2873;
Best Local Similarity 52.2%; Pred. No. 1.2e-07;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||     |  ||    ||     || ||      |     |   |   |
Db       1621 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 1680

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              |  | || | | || || |||||  | ||  | |||||  ||| |  | |||     |
Db       1681 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 1740

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |  |||||   ||||  || | |   ||   ||| || |||||||  |      || |||
Db       1741 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 1794

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||   |    |  |||||  |    | |||||     |   ||| | ||| |
Db       1795 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 1854

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  ||  |||||  ||||   ||||       |     | | |  | |
Db       1855 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 1914

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||   |  | ||   |  | || ||| || |||     ||| |||
Db       1915 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1974

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
               |  ||      |   | |   | ||||  |||  |  |       || |||   |
Db       1975 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 2034

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||    |||| ||| |  ||| | ||||||  |  |  |||||||  |  |
Db       2035 TGTGAAGAATGCGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGTACATT 2094

Qy        484 CATCATGGTG 493
              |||  ||| |
Db       2095 CATACTGGAG 2104
```

RESULT 6
US-10-104-047-1778
; Sequence 1778, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1778
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1778

Query Match 7.4%; Score 91.8; DB 17; Length 2110;
Best Local Similarity 51.7%; Pred. No. 1.1e-07;
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

```
Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              | | || || || |     | || || | |||  ||||  ||||| | ||| |   | |||
Db       1253 CATACCGAAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAAACACTCCTCA 1312

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
                  |      ||||   ||||  |||| |   ||   ||| || ||||||      ||
Db       1313 GCCCTTACTACACATAAGAGAATTCACACTGGAGAGAAACCCTACAAATGT------GAA 1366

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              || ||||  ||||| ||| | ||   ||| |  |     |||||    |  ||| |
Db       1367 GAATGTGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGAAACTTCATACT 1426

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              | ||| ||||| | | | ||| ||  ||||| ||||  |||           | | |
Db       1427 GGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTATCCAATCCTCAAAACTT 1486

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              |    |||| ||| | ||| |          |   | | | |||| | |||    ||
Db       1487 ACTGAACATAAAAAAATTCATTCTGGAGAGATACCCTACAAGTGTGAAGAATGTGGCAAA 1546

Qy        361 GCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAAAACA 417
              || ||| |  |      ||| |   | |||| | | | |   |  |  || ||| |
Db       1547 GCTTTTAAACACTCCTCATCCCTTACTACACATAAAAGAATTCATACTGGGGAGAAACCC 1606

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
              |  |  ||| || |||||   |||| ||| |  ||| | |||||| |  |  |||||||
Db       1607 TACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGATCCTCAAAACTTACTGAACATAAG 1666

Qy        478 TTAAAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
               |||  |||  ||| | ||   |||     ||    | |||     |      ||| |  ||
Db       1667 ATAATTCATACTGGAGAGAAACCCTATAAATGTGAGAGATGTGACAAAGCTTTTAACCAA 1726

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
              | | ||||| |      || || |  |  | | | |   |||| |   |||     | |
Db       1727 TCTGCAAACCT------TACTAAACATAAAAAAATACATACTGGAGAGAAACTACAGAAC 1780

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
              |   |  |||| ||| | || ||      |   | ||| |||  | ||||    || |||
Db       1781 TGGAATGTGTGATAATGATTTTGACAACACCTCAAATTTTTCTAAATATAAAGGAAATCA 1840

Qy        657 TGATGAT 663
              |  || |
Db       1841 TACTGGT 1847
```

RESULT 7
US-10-104-047-622
; Sequence 622, Application US/10104047
; Publication No. US20030236392A1

```
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
;   LENGTH: 3078
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-622

  Query Match             7.4%;  Score 91.8;  DB 17;  Length 3078;
  Best Local Similarity   52.1%;  Pred. No. 1.3e-07;
  Matches  257;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy         11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
              ||| | || |     |  ||     ||   | || ||      |          | | | | |
Db        650 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 709

Qy         71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
              |||||   || || || |||||  |||| |||||||| |||  |  | |||     |
Db        710 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 769

Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
              |||||    | ||  || | |    ||   ||| || ||||||       || || ||||  |
Db        770 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 823

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
              |||| |||  | |    ||| || |    | |||||   ||   ||| | | | | ||||
Db        824 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 883

Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | | |||  ||  ||||| ||||   ||||          |  | | |  | ||||
Db        884 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 943

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  |||||| |          |  | || |||| | |||    | || ||| |
Db        944 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 1003

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
              |       | |    | |||| | | | |   | |  | || ||| | |  |  |||
Db       1004 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 1063

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
              || |||||   ||||  || |    |   ||||   |  | |||||||  |||  |||
Db       1064 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 1123

Qy        488 ATGGTGGATCTCC 500
               ||| | |   ||
Db       1124 CTGGGGAAAAACC 1136


RESULT 8
US-10-094-749-692
; Sequence 692, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 692
;   LENGTH: 2114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-692

  Query Match             7.3%;  Score 90.4;  DB 17;  Length 2114;
  Best Local Similarity   51.8%;  Pred. No. 2.1e-07;
  Matches  258;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||    |      || ||      |  || |   |      |   | | |
Db       1024 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1083

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |   | || || |||||  |||| |||||| | ||| |  || |||     |
Db       1084 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 1143

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                ||||       |  ||   ||| ||   ||| || ||||||       || || |||
Db       1144 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1197

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |     | |||||   ||   ||||| | | |
Db       1198 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1257

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||  |   ||||| ||||| |||| |       |  | | | | | |
Db       1258 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1317

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  |  |||||              | | || |||| | |||    |||| |||
Db       1318 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1377

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||    |   ||   |  | |||| | ||  |   |  |   || ||| | |  |
Db       1378 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1437

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |  |   | ||||| |  |   ||||||  |||
Db       1438 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1497

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |      |||
Db       1498 CATACTGGAGAGAAACCT 1515


RESULT 9
US-10-094-749-499
; Sequence 499, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
```

```
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 499
;   LENGTH: 2230
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-094-749-499

  Query Match             7.3%;  Score 90.2;  DB 17;  Length 2230;
  Best Local Similarity   50.4%;  Pred. No. 2.4e-07;
  Matches  332;  Conservative    0;  Mismatches  313;  Indels   14;  Gaps    4;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||||| || |    |  ||    ||     || ||     |     |   | | |
Db        656 GAATGTGACAAATCACTTTGCATGCTTTCACGCCTAACTCAACATAAAAAAATTCATACT 715

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | | || || | | || || ||||| |||||  ||| ||| |||  |  |  ||   |
Db        716 AGAGAGAATTTCTACAAATGTGAAGAGTGTGGAAAAACCTTTAACTGGTCCACAAACCTT 775

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              ||| |   | ||  || | |   ||   ||| || ||||||  |||       |||
Db        776 TCTAAACCTAAGAAAATTCATACTGGAGAAAAACCCTACAAATGTGAAGTA------TGT 829

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |  |   |||    |    | |||||   ||   | | | | | ||
Db        830 GGAAAAGCCTTTCACCAATCCTCAATCCTTACTAAACATAAGATAATTCGTACTGGAGAA 889

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  |   | |||  ||  ||||| ||||   ||||          ||||  | |  |||
Db        890 AAACCCTATAAATGTGCACACTGTGGCAAAGCCTTTAAACAGTCCTCACACCTTACTAGA 949

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  | || |  |||||  |           |  | || |||| | |||    | |  |||
Db        950 CATAAGATAATTCATACTGAAGAGAAACCCTACAAATGTGAACAATGTGGCAAGGTCTTT 1009

Qy        367 TATAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |    | | |   |  | |||||| ||| | | |  |   ||  || ||| |
Db       1010 AAGCAGTCCCCAACCCTTACTAAACATCAGATAATTTATACTGGAGAGGAACCATACAAA 1069

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| |  |||||   |||| ||| |  | | ||| | || |  |  |||||||   ||
Db       1070 TGTGAGGAATGTGGCAAAGCTTTTAACCTATCTTAACAACTTACTGAACATAAGAAAATT 1129

Qy        484 CATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCA 542
              |   | | ||   ||| |  ||       ||     |      ||| |  | ||| ||
Db       1130 TACACTAGAGAGAAAGCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTTTTCA 1189

Qy        543 AACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAA 602
              | | |  | |    ||  ||     |||  | |     | ||    |  ||||||    |
Db       1190 ACCCTTATTACACATAAGATAATTCATAGCAGAGAGAAACCC----CACAAATGTGAAGA 1245

Qy        603 ATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATG 661
              |||||| | || || |     |      |     |  ||  ||||| ||||| ||||  ||
Db       1246 ATGTGGCAGAGCTTTTAACCAGTCCGCAAAGCTCACTGAACATAAGTTAATTCATACTG 1304


RESULT 10
US-10-029-386-13298
; Sequence 13298, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13298
;   LENGTH: 590
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR19.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
;   OTHER INFORMATION: NT HIT: gi14755315, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALUE 1.00e-119
;   OTHER INFORMATION: EST_HUMAN HIT: BG619358.1, EVALUE 0.00e+00
US-10-029-386-13298

  Query Match             7.2%;  Score 89.8;  DB 16;  Length 590;
  Best Local Similarity   53.8%;  Pred. No. 1.7e-07;
  Matches  234;  Conservative    0;  Mismatches  192;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
              || | ||| || | || | ||  | |||||  ||||  ||||| | |    |  | |||
Db        126 GTGCTAAAGAGAAAATCTACAAGTGTGAAGAATGTGGCAAAGCATTTCTATGGTCCTCAA 185

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
              ||     ||||   || |  |||| |    ||   ||| || ||||||      || |
Db        186 CCCTAACTAGACATAAGAGGATACACACTGGAGAGAAACCCTACAAATGT------GAAG 239

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
              | ||||  ||||| |||  |       |||    | |  | |||||    |  ||| | |
Db        240 AATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTG 299

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
              | | ||||| | | | |||  ||  ||||| ||||   |||    |   |  | |
Db        300 GAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGTTCTTCAACCCTTG 359

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
              | |||| | |  |  |||||               |  | || | || | |||    |||
Db        360 CTAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAG 419

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAA---AAAACAT 418
              | |||   ||      | |  |      ||||||   || ||  |  ||||   ||| | |
Db        420 CTTTTAGCAATTCCTCAACCCTTGCTAATCATAAGATAACTCATACTGAAGAGAAACCCT 479

Qy        419 TAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAAT 478
              |  |||||| ||||| | |||| ||| |  ||||  ||||    |  |  ||||||||
Db        480 ACAAATGTAAAGAATGTGACAAAGCTTTTAAGCGACTCTCAACCCTTACTAAACATAAAA 539

Qy        479 TAAAACATCATGGTG 493
              ||| ||||  ||| |
Db        540 TAATACATGCTGGAG 554


RESULT 11
US-10-108-260A-29
```

```
; Sequence 29, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
;   LENGTH: 2221
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-29

  Query Match             7.2%;  Score 89.2;  DB 17;  Length 2221;
  Best Local Similarity   53.4%;  Pred. No. 3.7e-07;
  Matches  237;  Conservative    0;  Mismatches  198;  Indels    9;  Gaps    2;

Qy          61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
               | | ||  | || |   || || || | | |  ||||  ||||||| |||| ||   |||
Db         654 CATACCGGAGAGAAACCTTTCAGATGTAAGGAATGTGGCAAAGCCTTTAATAGAAGCTCA 713

Qy         121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
               |    |     || |||    ||||  || | |    ||   ||| |  || ||||       |
Db         714 TACGTTACTCAGCATCAGAGAATGCATACTGGAGAGAAACCTTTCAAGTGTA------AG 767

Qy         181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
               || ||||  ||||| |||  ||||   |||||  |     ||||     |  |||  |
Db         768 GAATGTGGCAAAGCTTTTAACAGAGCTTCACACCTTACTCAACATCAGAGAATCCATACT 827

Qy         241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
               | | | ||||  ||| | |||  ||  ||||| ||||   ||||      |   |||| |
Db         828 GGAGAGAAACACTTCAAATGTAAAGAATGTGGCAAAGCCTTTAACAGGGGCTCACACCTT 887

Qy         301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
               |   ||||| |||  | ||||||                |  | || |||| | |||    ||
Db         888 ACTCGACATCAAAGAATCCATACTGGAGAGAAGTCTTTCAAATGTGAAGAATGTGGCAAA 947

Qy         361 GCATTT---TATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACA 417
               || |||   |        ||   |     |||||| | | |   |  |   || |||
Db         948 GCTTTTATCTGGGGCTCACACCTTACTCAACATCAGAGAGTCCATACTGGAGAGAAATTC 1007

Qy         418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
               || |  |||||| |||||   |||| ||| ||  |    ||| |  |  | ||||||
Db        1008 TTCAAATGTAAAGAATGTGGCAAAGCTTTTACCAGGAGCTCACACCTTACTCAACATCAG 1067

Qy         478 TTAAAACATCATGGTGGATCTCCT 501
                  ||  |||  ||| |    |||
Db        1068 AGAATTCATACTGGAGAGAAACCT 1091


RESULT 12
US-10-172-118-910
; Sequence 910, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Chris
;  APPLICANT:  Van 't Veer, Laura
;  APPLICANT:  Van de Vijver, Marc
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-175-999
;  CURRENT APPLICATION NUMBER: US/10/172,118
;  CURRENT FILING DATE:  2002-06-14
;  PRIOR APPLICATION NUMBER: 60/380,770
;  PRIOR FILING DATE: 2002-05-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
;   LENGTH: 3839
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: NM_003430
;   DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-910

  Query Match             7.2%;  Score 89.2;  DB 17;  Length 3839;
  Best Local Similarity   53.7%;  Pred. No. 4.6e-07;
  Matches  234;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    2;

Qy          69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
               | || |    | || || |||||  ||||  ||||| | ||  |   | |||    | |
Db         964 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1023

Qy         129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                |||||   ||||  || | |   ||   ||| || ||||||       || || ||||
Db        1024 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1077

Qy         189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               |||||  |||  |        |||    | |  | |||||     |   ||| | | | | ||
Db        1078 CAAAGCTTTTAGCCATTCTTCAGCCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 1137

Qy         249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
               ||| | | | |||  ||  ||||| ||||   |||     |       |  | |    |  ||
Db        1138 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 1197

Qy         309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
               | | || || |||||  |          |  | || | || | ||| |  || | ||| |
Db        1198 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 1257

Qy         369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                  ||     | |   |   | |||| | ||| ||  || |   || |||   |  | ||
Db        1258 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1317

Qy         426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
               | || |||||   |||| ||| | |||| |||||||  |   |   ||||| || |  ||
Db        1318 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1377

Qy         486 TCATGGTGGATCTCCT 501
               |  ||| |     |||
Db        1378 TACTGGAGAGAAACCT 1393


RESULT 13
US-10-342-887-910
; Sequence 910, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 910
;   LENGTH: 3839
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-10-342-887-910

  Query Match             7.2%;  Score 89.2;  DB 17;  Length 3839;
  Best Local Similarity   53.7%;  Pred. No. 4.6e-07;
  Matches  234;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    | || || |||||  ||||  ||||| | ||  |   | |||    | |
Db        964 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   ||||  || | |   ||   ||| || ||||||      || || ||||
Db       1024 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               ||||| |||  |       |||    | |  | |||||    |   ||| | | | | ||
Db       1078 CAAAGCTTTTAGCCATTCTTCAGCCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  |||||  ||||   |||    |      |  | |    |  ||
Db       1138 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || || |||||  |            | | || | || | ||| |  || | ||| |
Db       1198 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAACTTTTAA 1257

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
               ||    | |   |  | |||| | ||| ||  || |   || |||   |  |   ||
Db       1258 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   |||| ||| | |||| |||||||  |  |   ||||| || |  ||
Db       1318 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1377

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| |     |||
Db       1378 TACTGGAGAGAAACCT 1393


RESULT 14
US-10-098-841-48
; Sequence 48, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Xu, Chongjun
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Qian, Xiaohong B.
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2
;  CURRENT APPLICATION NUMBER: US/10/098,841
;  CURRENT FILING DATE:  2002-03-13
;  PRIOR APPLICATION NUMBER: 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 331
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 48
;   LENGTH: 3502
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (154)..(2583)
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(3502)
;   OTHER INFORMATION: n = a,t,c or g
US-10-098-841-48

  Query Match             7.2%;  Score 89;  DB 13;  Length 3502;
  Best Local Similarity   51.7%;  Pred. No. 4.8e-07;
  Matches  256;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy         79 TGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTA 138
              | || || |||||  ||||  ||||||| |||  |  | |||     |     |||||
Db       2101 TACAAATGTGAAGAATGTGGCAAAGCCTTTAAGTGGTCCTCAACCCTTACTAAACATAAG 2160

Qy        139 AGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTT 198
              | ||  || | |   ||   ||| || ||||||      || || ||||  ||||| |||
Db       2161 ATAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGGCAAAGCTTTT 2214

Qy        199 TTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCAT 258
                  |||    |   |||||||   |    ||| | | | | ||||| | | |
Db       2215 AAACTGTCCTCAACCCTTTCTACACATAAGATTATTCATACTGGAGAGAAACCCTACAAA 2274

Qy        259 TGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACC 318
              |||  |   ||||| ||||   ||||     |    | || | |    |||| | | |
Db       2275 TGTGAAAAATGTGGCAAAGCTTTTAACCGACCCTCAAACCTTATTGAACATAAGAAAATT 2334

Qy        319 CATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT---TATAAACAT 375
              |||||              | | || |||| | |||     ||||||||   |||
Db       2335 CATACTGGAGAGCAACCCTACAAATGTGAAGAATGTGGCAAAGCATTTAACTATTCCTCA 2394

Qy        376 CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGT 435
              ||   |  | |||| | | | |   |  | | ||  || | |  |  |||||| |||||
Db       2395 CACCTTAATACACATAAGAGAATTCATACTAAAGAGCAACCCTACAAATGTAAAGAATGT 2454

Qy        436 AATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGA 495
                |||| ||||  | |  ||||||  |  |   ||||||   ||  |||  ||| |
Db       2455 GGCAAAGCTTTCAACCAATATTCAAACCTTACTACACATAACAAAATTCATACTGGAGAG 2514

Qy        496 TCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTA 555
                 |     | |    |||    | |    ||| | ||     |||||||  |||
Db       2515 AAACTCTACAAACCTGAAGATGTGACAGTGATTTTGACAACACCTCAAACTTTTTCAAAC 2574

Qy        556 TTACAATTTCATATA 570
               || |||    ||||
Db       2575 ATAAAATAAATTATA 2589


RESULT 15
US-10-029-386-25074
; Sequence 25074, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25074
;   LENGTH: 1466
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC011467.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
;   OTHER INFORMATION: NT HIT: gi15309152, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P35789, EVALUE 0.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AU143734.1, EVALUE 0.00e+00
US-10-029-386-25074

  Query Match             7.2%;  Score 88.6;  DB 16;  Length 1466;
  Best Local Similarity   53.3%;  Pred. No. 4.1e-07;
  Matches  237;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |   |||| |   || || || | |||  ||||| ||  | | |      |  ||
Db        272 ACATACTGGAAAGAAATCTTTCAAATGTAAAGAATGTGAAAAGTCATTTTGCATGCTTTC 331

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
              |   ||||  |||||| |||||  || |||   ||   ||| || |||||||
Db        332 ACACTTAGCTCAACATAAAAGAATTCATAGTGGAGAGAAACCCTACAAATGTAAA----- 386

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
               || ||||  ||||| | |      |  ||| |  |    |||||||    |   ||| |
Db        387 -GAATGTGGGAAAGCCTATAATGAGACCTCAAACCTTTCTACACATAAAAGAATTCATAC 445

Qy        240 CGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
              | ||| ||||| | | | ||   || |||||| ||||   ||||    |    |||| |
Db        446 TGGAAAGAAACCCTACAAATGCGAAGAGTGTGGAAAAGCCTTTAACCGGCTCTCACACCT 505

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
              |  | |||| | || |  |||||              |  | || |||  | |||    |
Db        506 TACTACACATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAGGAGTGTGGCAA 565

Qy        360 AGCATTTTATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
              |||.||| |  ||  |    |   |   | |||| | | | |   |  |   || ||| |
Db        566 AGCTTTTAACCAATCTGCAAACCTTACTACACATAAGAGAATTCATACTGGAGAGAAACC 625

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
              |  |  ||| || |||||   | || ||| |   |   | ||||   |  |   ||||||
Db        626 CTACAAATGTGAAGAATGTGGCAGAGCTTTTAGCCAGTCCTCAACCCTTACTGCACATAA 685

Qy        477 ATTAAAACATCATGGTGGATCTCCT 501
                |||  |||  ||| |     |||
Db        686 GATAATTCATGCTGGAGAGAAACCT 710


Search completed: February  9, 2005, 11:50:32
Job time : 2162.11 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 01:45:39 ; Search time 3815.62 Seconds
(without alignments)
12360.150 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	182.2	14.7	628	8	BZ298000	BZ298000 CG3957.f1
c	2	167.6	13.5	959	9	CNS06MMP	AL405671 T7 end of
	3	103.8	8.4	801	7	CN759437	CN759437 ID0AAA25C
	4	103.8	8.4	898	7	CN761003	CN761003 ID0AAA2AE
	5	98.2	7.9	1609	3	AK032220	AK032220 Mus muscu
	6	97.8	7.9	2647	3	BC047646	BC047646 Homo sapi
	7	97.6	7.9	745	1	AU123448	AU123448 AU123448
	8	93.4	7.5	727	7	CO434402	CO434402 UI-M-HX0-
	9	92.4	7.5	631	7	CN352904	CN352904 170006000
	10	92.4	7.5	2672	3	BC036394	BC036394 Homo sapi
	11	92	7.4	591	8	AQ005136	AQ005136 CIT-HSP-2
	12	92	7.4	687	7	CN362485	CN362485 170004241
	13	92	7.4	752	7	CN417582	CN417582 170004240
	14	91	7.3	593	8	B99387	B99387 CIT-HSP-228
	15	91	7.3	820	5	BX414627	BX414627 BX414627
	16	90.8	7.3	666	7	CN297781	CN297781 170005321
	17	90.4	7.3	2694	3	BC022527	BC022527 Homo sapi
	18	90.4	7.3	2698	3	BC037782	BC037782 Homo sapi
	19	90.2	7.3	2253	3	BC028252	BC028252 Mus muscu
	20	89.8	7.2	550	5	BP292533	BP292533 BP292533
	21	89.2	7.2	582	5	BP234906	BP234906 BP234906
	22	89.2	7.2	604	2	BE161630	BE161630 MR3-HT044
	23	89.2	7.2	638	7	CN341035	CN341035 170006000
	24	89.2	7.2	673	7	CN353930	CN353930 170006001
	25	89.2	7.2	3609	3	BC032590	BC032590 Homo sapi
	26	89	7.2	2856	3	HSM803730	AL832422 Homo sapi
	27	88.6	7.2	535	1	AL705393	AL705393 DKFZp686M
	28	88.6	7.2	612	5	BP238960	BP238960 BP238960
	29	88.6	7.2	2174	3	AK033001	AK033001 Mus muscu
	30	88.6	7.2	2330	3	BC037426	BC037426 Homo sapi
	31	88.6	7.2	2622	3	AK033958	AK033958 Mus muscu
	32	88.4	7.1	998	5	BX456765	BX456765 BX456765
	33	88	7.1	895	5	BU195416	BU195416 AGENCOURT
	34	87.8	7.1	608	7	CN266758	CN266758 170004554
	35	87.8	7.1	4251	3	CR749856	CR749856 Homo sapi
	36	87.6	7.1	744	5	BX437291	BX437291 BX437291
	37	87.2	7.0	579	8	AQ347265	AQ347265 RPCI11-11
c	38	87.2	7.0	687	8	AQ389266	AQ389266 RPCI11-15
	39	87.2	7.0	907	5	BQ423752	BQ423752 AGENCOURT
	40	87	7.0	660	6	CD770140	CD770140 AGENCOURT
	41	87	7.0	842	5	BU507408	BU507408 AGENCOURT
c	42	86.8	7.0	521	8	AQ475626	AQ475626 CITBI-E1-
	43	86.8	7.0	571	7	CN280110	CN280110 170006001
	44	86.8	7.0	705	8	AQ194282	AQ194282 RPCI11-61
c	45	86.8	7.0	860	8	AQ749175	AQ749175 HS_5575_A

ALIGNMENTS

RESULT 1
BZ298000/c
LOCUS BZ298000 628 bp DNA linear GSS 31-OCT-2002
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata genomic clone CG3957, genomic survey sequence.
ACCESSION BZ298000
VERSION BZ298000.1 GI:24440936
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 628)
AUTHORS Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), R10 (2003)
MEDLINE 22508158
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.
FEATURES Location/Qualifiers
source 1. .628
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG3957"
/clone_lib="Candida glabrata Random Genomic Library"
ORIGIN

Query Match 14.7%; Score 182.2; DB 8; Length 628;
Best Local Similarity 59.5%; Pred. No. 7.5e-26;
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

```
Qy         84 ATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAAC 143
              ||| ||    ||||| || |||| ||   | || || ||  |      ||||    |  ||
Db        625 ATACGATAATTGTGACAAGGCCTTTACAAGGCCGTCCTTGCTCACAGAACACCAGAATAC 566

Qy        144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAG 203
                  ||  | |   | |    || | |    |  | ||||| |   |||||   |
```

```
Db        565 ---AGTGCATCTTGGGAGGAAACCTTGGAAATGTAATCAATGTGAAAGTTCATTTACTAA 509

Qy        204 AAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTC 263
              |||    || || || | |||  |      ||||  | ||  ||| ||| ||  |||||||
Db        508 AAAGATCCACTTAGAGAGACACTTATACACACACACAGATGAAAGACCGTTTTATTGTTC 449

Qy        264 AGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATAC 323
              | | ||||| || ||| | | | || | |||||  ||||| |||||||  | || || ||
Db        448 ATTTTGTGGGAAGGGGCTAATTACTAGGCAACAACTGAAACGACATGAGGTAACTCACAC 389

Qy        324 AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTT 383
              || || || || |||  || ||||  |||| | ||    || || || || |     ||
Db        388 CAAATCGTTCAATTGTGAATATGAAGGTTGTAATGAGAGTTTCTACAAGCACCCTCAATT 329

Qy        384 AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGT 443
              |||   |||||| ||| | |||||||  | |||   |||  ||| |  |||| || |||
Db        328 AAGGGCTCATATTTTAGCAGTTCATTTACAAAGTCTAAAATGTCACGAATGCAACAAAAG 269

Qy        444 TTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCAT------GGTGGATC 497
              |||     ||||||  | | |   | | || | |   |||| |||      | ||
Db        268 TTTTCAAAGACCTTACAGACTCAAAAATCACATAGCCAAACACCATAATCCAGATGTTGT 209

Qy        498 TCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATT 557
              |||||||||||||        || ||||| ||| | |||| |||| |||||  |   ||
Db        208 CAATGCTTATCAATGTACCTTCAGTGTTTGTTCTAAGAGTTTCAAAACATGGTCTGCGTT 149

Qy        558 ACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTG 617
              |  ||| ||| | ||| |    ||||| ||  | ||||| |||| |||  |||||    ||
Db        148 AAGATTACATGTTAAAAATGATCATCCGAAGTTAAAATGCCCTATATGCAGTAAACCCTG 89

Qy        618 TGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||||||| || | ||||||  |   || |||  ||  || |||||
Db         88 TGTTGGGGAAGACGGTTTAAATATGCACATGAAAATCCACGATGA 44
```

RESULT 2
CNS06MMP/c

```
LOCUS       CNS06MMP                 959 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AU0AA005F10 of library AU0AA from strain CBS 3082
            of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION   AL405671
VERSION     AL405671.1  GI:12168715
KEYWORDS    GSS.
SOURCE      Saccharomyces kluyveri
  ORGANISM  Saccharomyces kluyveri
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 959)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 959)
  AUTHORS   Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
            Gaillardin,C. and Casaregola,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 9.
            Saccharomyces kluyveri
  JOURNAL   FEBS Lett. 487 (1), 56-60 (2000)
  MEDLINE   20584719
   PUBMED   11152884
REFERENCE   3  (bases 1 to 959)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .959
                     /organism="Saccharomyces kluyveri"
                     /mol_type="genomic DNA"
                     /strain="CBS 3082"
                     /db_xref="taxon:4934"
                     /clone="AU0AA005F10"
                     /clone_lib="AU0AA"
                     /note="end : T7"
     misc_feature    complement(<71. .>958)
                     /note="similar to Saccharomyces cerevisiae ORF YPR186c [
                     PZF1 ; TFIIIA (transcription initiation factor) ]"
                     /evidence=not_experimental
ORIGIN
```

```
  Query Match             13.5%;  Score 167.6;  DB 9;  Length 959;
  Best Local Similarity   57.2%;  Pred. No. 6.1e-23;
  Matches  324;  Conservative    1;  Mismatches  235;  Indels    6;  Gaps    1;

Qy        211 CATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGT 270
              || ||||||| |||||  |  | |||||||| || |||||||| ||||||:|    |||
Db        958 CACTTGGAAAGACATATGTTCACGCATTCCGATAATAAACCATTTCATTGTYCCTATTGT 899

Qy        271 GGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCA 330
              |||||||| |||| | |    || || || |||||||||||| ||||| ||||| || ||
Db        898 GGTAAAGGTGTTACTACAAAGCAGCAGTTAAAAAGACATGAGATCACACATACCAAATCG 839

Qy        331 TTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACAT 390
              || ||||||  || |||    ||  | |||   ||||| || |||       ||||
Db        838 TTCAAATGTGAATATGAGGGCTGCAACGAATGTTTTTACAAGCATTCTCAACTAAGGTCA 779

Qy        391 CATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACT 450
              |||| | | || || ||      | | ||| || ||| | || ||   |||| | |||
Db        778 CATACACTTTCGGTACACTTGCAGAAATTGACCTGTGAGCATTGCGGTAAAATGTTCCAA 719

Qy        451 CGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCT------ 504
              | || |      |   | | |||     ||| |||||   ||    |  ||
Db        718 AGGCCATATCGTCTGCAAAACCATCTTGCAAAGCATCACAATGTGGATGTTGAAAACAAA 659

Qy        505 TATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTT 564
              |||||||||  | |  | |  ||      | | |||  |||| ||| | |    | ||
Db        658 TATCAATGTACTTACACGGCGTGCACAGAGACTTTTAAAACCTGGACTGCTCTTCAGCAG 599

Qy        565 CATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGG 624
              ||||| ||       ||||| ||  |  | |||      |||||| |||| |||||||||
Db        598 CATATCAAGGCGGACCATCCCAAGTTGCACTGTAACGTATGTGGCAAAGCTTGTGTTGGT 539

Qy        625 AAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGG 684
              |     ||| ||       || |||  |   ||||| ||  ||    | || ||||  |||
Db        538 GAGTCTGGTCTACAGATGCACATGCGAGTCCATGACGAGGCTTTAGTAATAAAAAATTGG 479

Qy        685 ACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTAT 744
              |  |||      |||| |    |      ||||| |||||   |||  |  |    || ||
Db        478 AAATGTACCATTTGTCAGCAAGTCTCTTTTGCTAAGAAGGCTGATCTTTTATCTCACTAC 419

Qy        745 AATATCTTCCATGATGGTAATATCCC 770
              |       ||| | | | |||| ||
Db        418 ATGGAACATCATAAGGATGATATACC 393
```

RESULT 3
CN759437

```
LOCUS       CN759437                 801 bp    mRNA    linear   EST 20-MAY-2004
```

```
DEFINITION  ID0AAA25CC01RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA25CC01
            5', mRNA sequence.
ACCESSION   CN759437
VERSION     CN759437.1  GI:47533360
KEYWORDS    EST.
SOURCE      Acyrthosiphon pisum (pea aphid)
  ORGANISM  Acyrthosiphon pisum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
            Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,
            Stern,D., Tagu,D. and Wincker,P.
  TITLE     An expressed sequence tags database for the pea aphid Acyrthosiphon
            pisum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: D. Tagu
            INRA Rennes
            UMR BiO3P, BP 35327, F-35653 Le Rheu Cedex France
            Tel: +33.2.23.48.51.65
            Fax: +33.2.23.48.51.50
            Risk of contamination by bacterial sequences from obligatory
            (Buchnera) or facultative endosymbionts. These sequences were
            obtained in the frame of the International Consortium of Aphid
            Genomics in collaboration with Genoscope
            PCR PRimers
            FORWARD: CAGGAAACAGCTATGACC
            Plate: 25  row: C  column: 1.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Acyrthosiphon pisum"
                     /mol_type="mRNA"
                     /cultivar="developmentstage"
                     /db_xref="taxon:7029"
                     /clone="ID0AAA25CC01"
                     /tissue_type="whole insect"
                     /dev_stage="nymphs and adults (parthenogenetic females)"
                     /lab_host="XL1-Blue"
                     /clone_lib="ApMS"
                     /note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI;
                     Sample name: ID0AAA ; Plant growth place: Department of
                     Ecology & Evolutionary Biology, Princeton University ;
                     Soil conditions: Soil ; Sowing date: 01/06/1999 ;
                     Harvesting date: 01/06/1999 ; Stress date: no stress ;
                     Description: Aphids inoculated on one-week old Vicia faba
                     under non-sterile conditions.  All parthenogenetic stages
                     and both winged and wingless adults were collected for
                     library construction. ; experimental condition: long
                     photoperiod (16-hr light/8-hr dark at 18 c)"
ORIGIN

  Query Match             8.4%;  Score 103.8;  DB 7;  Length 801;
  Best Local Similarity   51.5%;  Pred. No. 2.7e-10;
  Matches  318;  Conservative    0;  Mismatches  287;  Indels   12;  Gaps    3;

Qy         82 ACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGA 141
              | || |||    |||||||||| | || |  |  ||||   |||      |||  ||
Db          8 AAATGTGATTTCTGTGATAAAGAGTTTACTATAATATCACGTTTAACAAGGCATACAAAG 67

Qy        142 ACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTC 201
              || || |    |||    |||||| |||||      ||  | |||||| |||| ||||
Db         68 ACACATACCGGTGAAAAGCCGTATGAATGT------GATAACTGTGATCAAGCGTTTTCT 121

Qy        202 AGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGT 261
                |||||||||||    |   ||||   |  ||||| |||   |||| || |   | |||
Db        122 TTGAAATCACATTTAATAGGCCATAAAATGACACATACCGGCGAAAAGCCTTATAAATGT 181

Qy        262 TCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCAT 321
                |||| | ||| | ||  |     |  | | || | ||| |||  ||  || |||
Db        182 GATAACTGTGATCAAGCGTTTTCTAAGAAATCAAATTTAACAAGGCATATAAGGACACAT 241

Qy        322 ACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCT 381
              ||       |       | || || |||||| ||| |  |||| |||| | |  |   |  |

Db        242 ACCGGCGAAAAGCCGTATAAATGTGAAAAGTGTAACCAAGCGTTTTCTCAGAAATCAGAT 301

Qy        382 TTAAGACATCATATATTATCTGTTCAT---GAAAAAACATTAACGTGTAAACAATGTAAT 438
              |||| || |||||||    |   |      |||||  | |  |  ||| |  | ||| ||
Db        302 TTAATACGTCATATAAGGACACATACCGGCGAAAAGCCGTATAAATGTGATAACTGTGAT 361

Qy        439 AAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCT 498
               |||  ||  ||| |   |  | |||       |||||||| | ||||   |||| |
Db        362 CAAGCGTTTTCTCAAAAATGTAGTTTAATTAGCCATAAATTGACACATACCGGTGAAAAG 421

Qy        499 CCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTA 558
              |   | | | |||||| | |   || |          ||    ||       | |  |
Db        422 C---CGTTTAAATGTGCTAACTGTGATCAAGCGTTTTATCGGAAAGACCATTTAAAAAGC 478

Qy        559 CAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGT 618
              || |     | | |   |  ||         |||||||  || ||||    |||  | |
Db        479 CACTCAAGGACACATACAGGCCAAAAGCCGTATAAATGTGATACATGTAACCAAGCGTTT 538

Qy        619 GTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAA 678
                |  ||||  |  |||| |    ||||     |  ||||  | |   |   || | |||
Db        539 TCTCAGAAATCAAATTTAACAAGCCATACAAGGACACATGCCGGTGAAAAGCTGTTTAAA 598

Qy        679 ATATGGACTTGTGATTA 695
                  |  |||||| |
Db        599 TGCGATAACTGTGATAA 615


RESULT 4
CN761003
LOCUS       CN761003                 898 bp    mRNA    linear   EST 20-MAY-2004
DEFINITION  ID0AAA2AE05RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA2AE05 5',
            mRNA sequence.
ACCESSION   CN761003
VERSION     CN761003.1  GI:47534926
KEYWORDS    EST.
SOURCE      Acyrthosiphon pisum (pea aphid)
  ORGANISM  Acyrthosiphon pisum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
            Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,
            Stern,D., Tagu,D. and Wincker,P.
  TITLE     An expressed sequence tags database for the pea aphid Acyrthosiphon
            pisum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: D. Tagu
            INRA Rennes
            UMR BiO3P, BP 35327, F-35653 Le Rheu Cedex France
            Tel: +33.2.23.48.51.65
            Fax: +33.2.23.48.51.50
            Risk of contamination by bacterial sequences from obligatory
            (Buchnera) or facultative endosymbionts. These sequences were
            obtained in the frame of the International Consortium of Aphid
            Genomics in collaboration with Genoscope
            PCR PRimers
            FORWARD: CAGGAAACAGCTATGACC
            Plate: 2  row: E  column: 5.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Acyrthosiphon pisum"
                     /mol_type="mRNA"
                     /cultivar="developmentstage"
                     /db_xref="taxon:7029"
                     /clone="ID0AAA2AE05"
                     /tissue_type="whole insect"
                     /dev_stage="nymphs and adults (parthenogenetic females)"
                     /lab_host="XL1-Blue"
                     /clone_lib="ApMS"
                     /note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI;
                     Sample name: ID0AAA ; Plant growth place: Department of
                     Ecology & Evolutionary Biology, Princeton University ;
                     Soil conditions: Soil ; Sowing date: 01/06/1999 ;
```

Harvesting date: 01/06/1999 ; Stress date: no stress ;
Description: Aphids inoculated on one-week old Vicia faba under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 c)"

ORIGIN

Query Match 8.4%; Score 103.8; DB 7; Length 898;
Best Local Similarity 51.5%; Pred. No. 2.7e-10;
Matches 318; Conservative 0; Mismatches 287; Indels 12; Gaps 3;

```
Qy        82 ACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGA 141
             | || |||    |||||||||| | || | |  ||||  |||      |||  ||
Db         7 AAATGTGATTTCTGTGATAAAGAGTTTACTATAATATCACGTTTAACAAGGCATACAAAG 66

Qy       142 ACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTC 201
             || || |    |||    |||||| |||||      ||  | ||||||| |||| ||||
Db        67 ACACATACCGGTGAAAAGCCGTATGAATGT------GATAACTGTGATCAAGCGTTTTCT 120

Qy       202 AGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGT 261
             ||||||||||    |  ||||   |  ||||| |||   |||| || |   | |||
Db       121 TTGAAATCACATTTAATAGGCCATAAAATGACACATACCGGCGAAAAGCCTTATAAATGT 180

Qy       262 TCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCAT 321
             ||||  | ||| | ||  |     |  | | || | ||| |||  ||  || |||
Db       181 GATAACTGTGATCAAGCGTTTTCTAAGAAATCAAATTTAACAAGGCATATAAGGACACAT 240

Qy       322 ACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCT 381
             ||      |      | || || |||||| ||| |  |||| |||| | |  |   |  |
Db       241 ACCGGCGAAAAGCCGTATAAATGTGAAAAGTGTAACCAAGCGTTTTCTCAGAAATCAGAT 300

Qy       382 TTAAGACATCATATATTATCTGTTCAT---GAAAAAACATTAACGTGTAAACAATGTAAT 438
             |||| || |||||||    |   |      |||||  | |  |  ||| |  | ||| ||
Db       301 TTAATACGTCATATAAGGACACATACCGGCGAAAAGCCGTATAAATGTGATAACTGTGAT 360

Qy       439 AAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCT 498
              |||  ||  ||| |   |  |  |||      |||||||| | ||||   |||| |
Db       361 CAAGCGTTTTCTCAAAAATGTAGTTTAATTAGCCATAAATTGACACATACCGGTGAAAAG 420

Qy       499 CCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTA 558
             |   | | | ||||||| | |   || |           ||     ||       | | |
Db       421 C---CGTTTAAATGTGCTAACTGTGATCAAGCGTTTTATCGGAAAGACCATTTAAAAAGC 477

Qy       559 CAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGT 618
             || |     | | |   |   ||          |||||||  || ||||    |||  | |
Db       478 CACTCAAGGACACATACAGGCCAAAAGCCGTATAAATGTGATACATGTAACCAAGCGTTT 537

Qy       619 GTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAA 678
              |  ||||  |  |||| |    ||||     |  ||||  | |   |    || | |||
Db       538 TCTCAGAAATCAAATTTAACAAGCCATACAAGGACACATGCCGGTGAAAAGCTGTTTAAA 597

Qy       679 ATATGGACTTGTGATTA 695
                 |  |||||| |
Db       598 TGCGATAACTGTGATAA 614
```

RESULT 5
AK032220
LOCUS AK032220 1609 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430503O15 product:MSZF33 (FRAGMENT) homolog [Mus musculus], full insert sequence.
ACCESSION AK032220
VERSION AK032220.1 GI:26328050
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1609)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6430503O15"
/db_xref="taxon:10090"
/clone="6430503O15"
/sex="male"

```
                     /tissue_type="olfactory brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             330. .1223
                     /note="unnamed protein product; MSZF33 (FRAGMENT) homolog
                     [Mus musculus] (SPTR|O88251, evidence: FASTY, 87.1%ID,
                     100%length, match=255)
                     putative"
                     /codon_start=1
                     /protein_id="BAC27766.1"
                     /db_xref="GI:26328051"
                     /translation="MLETYRNLNAVGYNWEDSNIEEHCESSRRHGRHERNHTGEKPYE
                     GIQYGEAFVHHSSLQMRKIIHTGEKRYKCNQCDKAYSRHSILQIHKRTHSGEKPYECN
                     QCGKAFTQHSHLKIHMVTHTGEKPYKCNQCGKAFACHNKLQKHERTHTGEKPYKCDQC
                     NKAFVYESYLQVHKKTHTGEKPYKCNECGKAFARHSHLKVHKITHTGEKPYKCNQCGK
                     ALAYHSTLQVHQRTHTGEKPYECEQCGKAFANQSYFQVHKRIHTGEKPYKCDQCGKAF
                     VGSSDLKRHERVHTGRETLQM"
     polyA_signal    1594. .1599
                     /note="putative"
     polyA_site      1609
                     /note="putative"
ORIGIN

  Query Match             7.9%;  Score 98.2;  DB 3;  Length 1609;
  Best Local Similarity   51.2%;  Pred. No. 3.6e-09;
  Matches  310;  Conservative    0;  Mismatches  283;  Indels   12;  Gaps    3;

Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |    | || |     | || ||    |    ||||||||||||||   ||||
Db        521 ACATACTGGAGAGAAACGCTACAAATGCAATCAATGTGATAAAGCCTATTCACGACACAG 580

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
                |  || |   ||||  |||||| || |||   ||   ||| ||| ||||||      |
Db        581 CATTCTACAAATACATAAAAGAACACATAGTGGAGAGAAACCCTATGAATGTA------A 634

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
              | |||| |||||| |||      | |    ||| |  ||| |||||| ||  ||||| |
Db        635 TCAATGTGGTAAAGCCTTTACACAACACAGTCATCTCAAAATACATATGGTTACACATAC 694

Qy        240 CGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
              | | | ||||| | | | |||       ||||||||||   ||   | ||   |  |  |
Db        695 TGGAGAGAAACCTTACAAATGTAATCAATGTGGTAAAGCCTTTGCATGTCATAATAAACT 754

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
                ||| ||||||||  || |||||              |  |  | |||  | ||| |  |
Db        755 TCAAAAACATGAAAGAACACATACTGGAGAGAAACCTTACAAGTGTGATCAATGTAATAA 814

Qy        360 AGCATTT---TATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
              |||  |||   |||  ||  | | |     |   ||| | | |  |  | |  |   || ||| |
Db        815 AGCCTTTGTATATGAAAGTTATTTACAAGTTCATAAAAAAACACATACTGGAGAGAAACC 874

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
              |  |  |||||  |||||  |||||  ||  | ||||       |  |    |  |||||
Db        875 TTACAAATGTAATGAATGTGGTAAAGCCTTTGCACGACACAGTCATCTCAAAGTGCATAA 934

Qy        477 ATTAAAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
              | ||| ||||  ||| |      |   ||||  |||||  ||||   ||||
Db        935 AATAACACATACTGGAGAGAAAC---CTTACAAATGTAATCAATGTGGTAAAGCCCTTGC 991

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
              |  || | |         | ||||       | | |         |         | ||||
Db        992 ATATCATAGTACACTCCAAGTACATCAAAGAACACATACTGGAGAGAAGCCCTATGAATG 1051

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
              |    ||||||| ||||  | ||      |||    |||      |||||    ||  ||
Db       1052 TGAGCAATGTGGCAAAGCCTTTGCAAATCAAAGTTATTTCCAAGTACATAAAAGAATACA 1111

Qy        657 TGATG 661
              |  ||
Db       1112 TACTG 1116


RESULT 6
```

```
BC047646
LOCUS       BC047646               2647 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone
            IMAGE:4828290), with apparent retained intron.
ACCESSION   BC047646
VERSION     BC047646.1  GI:29126804
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2647)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2647)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 90 Row: i Column: 7
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4828290"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN

  Query Match             7.9%;  Score 97.8;  DB 3;  Length 2647;
```

```
  Best Local Similarity   50.1%;  Pred. No. 4.5e-09;
  Matches  330;  Conservative    0;  Mismatches  317;  Indels   12;  Gaps    3;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | ||||         ||  ||      |  || |   |     |  | | |
Db       1235 GAATGTGGCAAAACTTTTAACCGATTCTCAACTCTTACTACCCATAAGATAATTCATACT 1294

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | ||  |    | || || | |||  |||| |||||| | ||| ||| | |||     |
Db       1295 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTT 1354

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                 |||  || ||  || | |    ||   ||| || ||||||      || || |||
Db       1355 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 1408

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||         ||| |  |    |||||||   ||   ||| | | | |
Db       1409 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 1468

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  | | | |||  |   |||||  ||||   ||||          |||| | |  | |
Db       1469 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACA 1528

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||||  | |  |||||                |  | || |||||| |||  | |||| |||
Db       1529 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 1588

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               || |     ||   |  |||| | ||| |   |  |  || ||| | |  |   |
Db       1589 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 1648

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              |||||| |||||  ||||| ||| |  |   ||||||   |  |  |||||||  |||
Db       1649 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 1708

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              |||  ||| |    |||    |      ||   |          ||||||  | | |  |
Db       1709 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 1768

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
              |      | | |     |   |  |  | | ||  |  | |||||    |||
Db       1769 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 1825

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              |||||  |||| || |      |      | | |  ||  |||||  | |||||||   ||
Db       1826 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1884


RESULT 7
AU123448
LOCUS       AU123448                745 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU123448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
            sequence.
ACCESSION   AU123448
VERSION     AU123448.1  GI:10948164
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
            Suzuki,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RM2000319"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RM2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
ORIGIN

  Query Match             7.9%;  Score 97.6;  DB 1;  Length 745;
  Best Local Similarity   50.2%;  Pred. No. 4.5e-09;
  Matches  328;  Conservative    0;  Mismatches  314;  Indels   12;  Gaps    3;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | ||||         ||| ||      |  || |   |     |  | | |
Db         87 GAATGTGGCAAAACTTTTAACCGATCCTCAACTCTTACTACCCATAAGATAATTCATACT 146

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | ||  |    | || || | |||  |||| |||||| | ||| ||  | |||     |
Db        147 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGTTCTTCAACCCTT 206

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                 |||  || ||  || | |    ||   ||| || ||||||      || || |||
Db        207 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 260

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||         ||| |  |    |||||||   ||   ||| | | | |
Db        261 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 320

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  | | | |||  |   |||||  ||||   ||||          |||| | |  | |
Db        321 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACTACA 380

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||||  | |  |||||                |  | || |||||| |||  | |||| |||
Db        381 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 440

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               || |     ||   |  |||| | ||| |   |  |  || ||| | |  |   |
Db        441 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 500

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              |||||| |||||  ||||| ||| |  |   ||||||   |  |  |||||||  |||
Db        501 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 560

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              |||  ||| |    |||    |      ||   |          ||||||  | | |  |
Db        561 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 620

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
              |      | | |     |   |  |  | | ||  |  | |||||    |||
Db        621 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 677

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCAT 657
              |||||  |||| || |      |      | | |  ||  |||||  | |||||||
Db        678 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCAT 731


RESULT 8
CO434402
LOCUS       CO434402                727 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  UI-M-HX0-csl-h-09-0-UI.r1 NIH_BMAP_HX0 Mus musculus cDNA clone
            IMAGE:30687440 5', mRNA sequence.
```

```
ACCESSION   CO434402
VERSION     CO434402.1  GI:49680696
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30687440"
                     /tissue_type="whole eye"
                     /dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16,
                     17, 18 dpc)"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_HX0"
                     /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AATAATTACG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH)."
ORIGIN
```

```
  Query Match             7.5%;  Score 93.4;  DB 7;  Length 727;
  Best Local Similarity   52.1%;  Pred. No. 3e-08;
  Matches  261;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          4 AGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGT 63
              | | |  ||||  || ||   ||   ||     |     |   |||  |     | ||| |
Db         78 AATCAGTGTGATAAAGCCTTTTCTCGATACAGTACTCTTCAATCTCATAGAAGAACACAT 137

Qy         64 CCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTA 123
                |  || |    | || || | |  |||||||||||||| |  |     ||
Db        138 ACTGGAGAGAAACCCTTCAAATGTAATCAGTGTGATAAAGCCTTTTCTGAGAAATGTAGT 197

Qy        124 TTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGAT 183
                |  |  | ||| |||| |  || | |   ||   ||| || |||||||      |  |
Db        198 CTCCAAAACCATGTAAGGATACATACTGGAGAGAAACCCTACAAATGTA------ATCAA 251

Qy        184 TGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAA 243
              |||||||||| |||| |   || |  ||  |  | |  ||||    | ||||| ||| |
Db        252 TGTGATAAAGCCTTTTCCCATAACTATCACCTCCACATTCATAGAAGAACACATACCGGA 311

Qy        244 AAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAA 303
                | ||||| | | |||     ||||| |||||   ||     |    | ||| |  |
Db        312 GAGAAACCCTACAAATGTAATCAGTGTGATAAAGCCTTTTCCCATAACTATCACCTCCAC 371

Qy        304 AGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCA 363
              |  |||  ||  || |||||               |  | ||| | |  | ||| |  ||||
Db        372 ATTCATAGAAGAACACATACTGGAGAGAAGCCTTACAAATGTAATCAGTGTGATAAAGCC 431

Qy        364 TTTTATAAACATCAATCTTTAAGACATCATATATTATC---TGTTCATGAAAAAACATTA 420
              |||| | || |    ||| |   |  ||| | |  | |   |  |   || ||| ||||
Db        432 TTTTCTCAATACTTTTCTCTTCAAACTCACAGAAGAACACATACTGGAGAGAAACCATTC 491

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              |  ||||| |||||| |||||||  ||  ||| |       |   |   |  |||| | |
Db        492 AAATGTAATCAATGTGATAAAGTCTTTTCTCAATACAGTACTCTTCAAACTCATAGAAGA 551

Qy        481 AAACATCATGGTGGATCTCCT 501
              | ||||  |||  |     |||
Db        552 ACACATACTGGAGAGAAACCT 572
```

```
RESULT 9
CN352904
LOCUS       CN352904                631 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000600057466 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN352904
VERSION     CN352904.1  GI:47352838
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 631   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, DMSO-treated H9 cell
                     line"
                     /clone_lib="GRN_PREHEP"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from DMSO-treated hES cell line H9 (p22) maintained in
                     feeder-free conditions"
ORIGIN
```

```
  Query Match             7.5%;  Score 92.4;  DB 7;  Length 631;
  Best Local Similarity   54.1%;  Pred. No. 4.7e-08;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |    || || || | |||  ||||| ||||| | ||| | | | |||   | |
Db         83 AGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAAACTTAC 142

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                |||||   | ||  || | |   ||   ||| ||| |||||       ||  | ||||
Db        143 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTATGAATGT------GAAAAATGTGG 196

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||  |||  |       ||| || |    | |||||     |   ||| | ||| | ||
Db        197 CAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGAAGAGAA 256
```

```
Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  ||||| |||||  |||| |  |    |  | | |  |  ||
Db        257 ACCTTACAAATGTGAAGAATGTGGCAAAGGTTTTAAATGGCCCTCAACCCTTACTATCCA 316

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || |  |||||             ||  | || |||| | |||    |||| ||| |
Db        317 TAAGATAATTCATACTGGAGAGAAACCATACAAATGTGAAGAATGTGGCAAAGCTTTTAA 376

Qy        369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              ||       ||  |   | |||| | | | |    |  |   || ||| | |  || ||
Db        377 CCAATCCTCAAAACTTACCAAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATG 436

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   ||||  || |   |   | |||||  |  |  |||||||   ||  ||
Db        437 TGAAGAATGTGGCAAAGCCTTTAACCAGTCCTCAAACCTTACTAAACATAAGAGAATTCA 496

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| | |   |||
Db        497 TACTGGAGAAAAACCT 512
```

RESULT 10
BC036394
LOCUS BC036394 2672 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone IMAGE:5259399), containing frame-shift errors.
ACCESSION BC036394
VERSION BC036394.1 GI:23025784
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2672)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2672)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 73 Row: a Column: 16
This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
source 1. .2672
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5259399"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
ORIGIN

Query Match 7.5%; Score 92.4; DB 3; Length 2672;
Best Local Similarity 54.1%; Pred. No. 5.3e-08;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   || || || | |||  ||||| ||||| | ||| | | | |||  | |
Db       1745 AGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAAACTTAC 1804

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   | ||  || | |   ||   ||| ||| |||||      ||  | ||||
Db       1805 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTATGAATGT------GAAAAATGTGG 1858

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
               ||||| ||| |       ||| || |    | |||||     |   ||| | ||| | ||
Db       1859 CAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGAAGAGAA 1918

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  ||||| |||||  |||| |  |    |  | | |  |  ||
Db       1919 ACCTTACAAATGTGAAGAATGTGGCAAAGGTTTTAAATGGCCCTCAACCCTTACTATCCA 1978

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || |  |||||             ||  | || |||| | |||    |||| ||| |
Db       1979 TAAGATAATTCATACTGGAGAGAAACCATACAAATGTGAAGAATGTGGCAAAGCTTTTAA 2038

Qy        369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              ||       ||  |   | |||| | | | |    |  |   || ||| | |  || ||
Db       2039 CCAATCCTCAAAACTTACCAAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATG 2098

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||   ||||  || |   |   | |||||  |  |  |||||||   ||  ||
Db       2099 TGAAGAATGTGGCAAAGCCTTTAACCAGTCCTCAAACCTTACTAAACATAAGAGAATTCA 2158

Qy        486 TCATGGTGGATCTCCT 501
              |  ||| | |   |||
Db       2159 TACTGGAGAAAAACCT 2174
```

RESULT 11
AQ005136
LOCUS AQ005136 591 bp DNA linear GSS 27-JUN-1998
DEFINITION CIT-HSP-2290O17.TF CIT-HSP Homo sapiens genomic clone 2290O17, genomic survey sequence.
ACCESSION AQ005136
VERSION AQ005136.1 GI:3082581
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 591)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

```
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2290O17.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7150844"
                     /db_xref="taxon:9606"
                     /clone="2290O17"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.4%;  Score 92;  DB 8;  Length 591;
  Best Local Similarity   52.2%;  Pred. No. 5.7e-08;
  Matches  256;  Conservative    0;  Mismatches  225;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||    |      || ||      |  || |   |     |  | | |
Db         69 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 128

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |    | || || |||||  |||| |||||| | ||| |  || |||    |
Db        129 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 188

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||        |  ||   ||| ||   ||| || ||||||      || || |||
Db        189 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 242

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |    | |||||   ||   ||||| | | |
Db        243 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 302

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||  |   ||||| |||||  |||| |       |  | | | || |
Db        303 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 362

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| |||   | |||||               | | || |||| | |||    |||| |||
Db        363 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 422

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               ||     |   ||   |  | |||| | ||  |   |  |   || ||| | |  |
Db        423 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 482

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| || |  |   | |||||| |  |   ||||||  |||
Db        483 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 542

Qy        484 CATCATGGTG 493
              |||  ||| |
Db        543 CATACTGGAG 552


RESULT 12
CN362485
```

```
LOCUS       CN362485                687 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000424188118 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN362485
VERSION     CN362485.1  GI:47362419
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 687   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, embryoid bodies
                     derived from H1, H7 and H9 cells"
                     /clone_lib="GRN_EB"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from embryoid body outgrowths derived from hES cell lines
                     H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                     conditions."
ORIGIN

  Query Match             7.4%;  Score 92;  DB 7;  Length 687;
  Best Local Similarity   52.2%;  Pred. No. 5.7e-08;
  Matches  256;  Conservative    0;  Mismatches  225;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||    ||      || ||      |      |   |  |
Db        183 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 242

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              |  |  | || | | || || |||||  | ||  | ||||| |||  |  | |||     |
Db        243 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 302

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  |||||   ||||  || | |   ||   ||| || |||||||  |       || |||
Db        303 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 356

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||   |   |  |||||   |    | |||||    |   ||| | ||| |
Db        357 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 416

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | |||  ||  |||||  ||||    ||||       |  | | |  | |
Db        417 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 476

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | || |  |||||    |  | ||    |  | || ||| || |||      ||| |||
Db        477 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 536

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
              |  ||      |   | |    | ||||  |||  |  |       || |||   |
Db        537 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAACGTTATGAA 596

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||    |||| ||| |  ||| | |||||| |  |  |||||||  |  |
Db        597 TGTGAAGAATGCGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGTACATT 656
```

```
Qy        484 CATCATGGTG 493
              ||| ||| |
Db        657 CATACTGGAG 666


RESULT 13
CN417582
LOCUS       CN417582               752 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000424010604 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN417582
VERSION     CN417582.1  GI:47405176
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 752   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, embryoid bodies
                     derived from H1, H7 and H9 cells"
                     /clone_lib="GRN_EB"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from embryoid body outgrowths derived from hES cell lines
                     H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                     conditions."
ORIGIN

  Query Match             7.4%;  Score 92;  DB 7;  Length 752;
  Best Local Similarity   52.0%;  Pred. No. 5.8e-08;
  Matches  259;  Conservative    0;  Mismatches  230;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |     || ||     | || |   |     |   | | |
Db         72 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 131

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |    | || || |||||  |||| |||||| | ||| |  || |||    |
Db        132 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 191

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |    ||||      |  ||   ||| ||   ||| || ||||||      || || |||
Db        192 GCTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 245

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| | |    | |||||   ||   ||||| | | |
Db        246 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 305

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||  |   ||||| |||||  |||| |       |  | | | || |
Db        306 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 365

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  | |||||              | | || |||| | |||    |||| |||
Db        366 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 425
```

```
Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||     |   ||   |   | |||| | ||  |   |  |   || ||| | | |
Db        426 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 485

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||   |||| || | |  | ||||||  |  |   ||||||  |||
Db        486 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 545

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |     |||
Db        546 CATACTGGAGAGAAACCT 563


RESULT 14
B99387
LOCUS       B99387                 593 bp    DNA     linear   GSS 26-JUN-1998
DEFINITION  CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
            genomic survey sequence.
ACCESSION   B99387
VERSION     B99387.1  GI:3027197
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7147044"
                     /db_xref="taxon:9606"
                     /clone="2281A10"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.3%;  Score 91;  DB 8;  Length 593;
  Best Local Similarity   54.4%;  Pred. No. 8.9e-08;
  Matches  231;  Conservative    0;  Mismatches  185;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              |||| |    | || || |||||  ||||  ||||| | ||  |    | |||    |
Db         47 AAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCAGTCCTCAACCCTTAG 106

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
              |||||   | ||  || | |   ||   ||| || ||||||      || || ||||
Db        107 AAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 160

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||| |||    |   |||||  |   | |||||  |||   ||| | ||| | ||
```

```
Db        161 TAAAGCTTTTAAGTGGTCCTCACACCTTACTAGACATAAAGTAATTCATACTGAAGAGAA 220

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308

Db        221 ACCCTACAAATGTGAAGAATGTGGCAAGGCTTTTAACCATTTCTCAGCCCTTAGGAAACA 280

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368

Db        281 TAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAG 340

Qy        369 TAAACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425

Db        341 CCAGTCCTCAACTCTTAGAAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATG 400

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485

Db        401 TGAAGAATGTGGTAAAGCTTTTAAGTGGTCCTCAAAACTTACTGTACATAAGGTAATTCA 460

Qy        486 TCATG 490

Db        461 TACTG 465


RESULT 15
BX414627
LOCUS       BX414627                820 bp    mRNA    linear   EST 01-MAY-2004
DEFINITION  BX414627 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YD12
            5-PRIME, mRNA sequence.
ACCESSION   BX414627
VERSION     BX414627.2  GI:46930382
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 13, 2003 this sequence version replaced gi:30649903.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 9703.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0CAP001DB06QP1&c=9703.f.
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YD12"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             7.3%;  Score 91;  DB 5;  Length 820;
  Best Local Similarity   50.1%;  Pred. No. 9.2e-08;
  Matches  289;  Conservative    8;  Mismatches  268;  Indels   12;  Gaps    3;

Qy         89 AAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACA 148

Db          2 AAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTACTACSCATAGARGRRTTCATA 61

Qy        149 GTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAAT 208

Db         62 CTGGAGAGAAACCTTACAAATGT------GAAGAATGKGGCAAAGCCTTTAAGCAGTCCT 115

Qy        209 CACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGT 268

Db        116 CAAACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAAGAT 175

Qy        269 GTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGT 328

Db        176 GTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACATGAGGTAATTCATACTGGAG 235

Qy        329 CATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACATCAATCTTTAA 385

Db        236 AGAAACCCTACAAATGTGAAAGATGTGGAAAAGCCTTTAATCATTTCTCACACCTTACTA 295

Qy        386 GACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTT 445

Db        296 CACATAAGATAATTCATACTGGAGAGAAACCTTACAAATGTAAAGRATGTGGTAAAGCTT 355

Qy        446 TCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTT 505

Db        356 TTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCATACTGGAGAGAAGCCTTACA 415

Qy        506 ATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTC 565

Db        416 AATGTAAAGRATGTGAAAGAGCTTTTAACCAATCCTCAAAACTTACTGAACATAAGARAA 475

Qy        566 ATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGA 625

Db        476 TTCATACTGGAGAGAAACCCTA---TGAATGTGAAARATGTGGCAAAGCTTTTAACCAGT 532

Qy        626 AAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662

Db        533 CCTCAAATCTTACTAGACATAAGAGRAGTCATACAGA 569


Search completed: February  9, 2005, 07:43:38
Job time : 3817.62 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:03:15 ; Search time 73 Seconds
(without alignments)
2182.812 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2229	100.0	412	3	AAY93316	Aay93316 A transcr
2	2225	99.8	412	5	ABP73656	Abp73656 Candida a
3	445	20.0	564	6	ABJ25655	Abj25655 Aspergill
4	445	20.0	564	6	ABJ26255	Abj26255 Aspergill
5	416.5	18.7	1212	4	ABG00399	Abg00399 Novel hum
6	407	18.3	664	8	ADN99509	Adn99509 Novel hum
7	403	18.1	1230	4	AAU30831	Aau30831 Novel hum
8	400	17.9	404	7	ADM05916	Adm05916 Human pro
9	400	17.9	803	8	ADN06034	Adn06034 Antipsori
10	400	17.9	809	4	AAM38689	Aam38689 Human pol
11	399	17.9	444	8	ADN99691	Adn99691 Novel hum
12	398	17.9	444	8	ADN99690	Adn99690 Novel hum
13	398	17.9	487	8	ABO60215	Abo60215 Human gen
14	398	17.9	563	6	ABU96690	Abu96690 Human nuc
15	398	17.9	563	7	ADI60246	Adi60246 Secreted
16	397.5	17.8	423	8	ADQ29657	Adq29657 Human col
17	397.5	17.8	423	8	ADP24748	Adp24748 PRO polyp
18	395.5	17.7	365	3	AAY93317	Aay93317 A human t
19	395.5	17.7	409	4	AAG75181	Aag75181 Human col
20	395.5	17.7	423	2	AAR91305	Aar91305 Transcrip
21	395.5	17.7	1191	7	ADJ70299	Adj70299 Human hea
22	395	17.7	622	4	AAM78947	Aam78947 Human pro
23	395	17.7	631	4	AAM79931	Aam79931 Human pro
24	395	17.7	632	4	ABG18386	Abg18386 Novel hum
25	395	17.7	719	4	ABG16953	Abg16953 Novel hum
26	393.5	17.7	1191	8	ADN04208	Adn04208 Antipsori
27	393.5	17.7	1472	4	AAU31578	Aau31578 Novel hum
28	393	17.6	517	4	ABG03375	Abg03375 Novel hum
29	392.5	17.6	1050	4	ABG09685	Abg09685 Novel hum
30	392.5	17.6	1050	7	ADC32995	Adc32995 Human nov
31	392	17.6	555	6	ADA54763	Ada54763 Human pro
32	391	17.5	592	7	ADM04984	Adm04984 Human pro
33	390.5	17.5	878	4	AAU29528	Aau29528 Novel hum
34	390.5	17.5	1520	4	ABG27130	Abg27130 Novel hum
35	390	17.5	540	7	ADB65594	Adb65594 Human pro
36	390	17.5	576	6	ADA54797	Ada54797 Human pro
37	390	17.5	620	6	ABU96725	Abu96725 Human nuc
38	389	17.5	891	8	ADQ67674	Adq67674 Novel hum
39	388.5	17.4	501	4	ABB71369	Abb71369 Drosophil
40	388	17.4	568	5	ABB79872	Abb79872 TRAF6-inh
41	387.5	17.4	723	8	ADN99696	Adn99696 Novel hum
42	387.5	17.4	742	8	ABO58351	Abo58351 Human gen
43	387.5	17.4	839	7	ADJ70044	Adj70044 Human hea
44	387.5	17.4	872	7	ADC37535	Adc37535 Human nuc
45	386.5	17.3	327	8	ADR08458	Adr08458 Human pro

ALIGNMENTS

RESULT 1
AAY93316
ID AAY93316 standard; protein; 412 AA.
XX
AC AAY93316;
XX
DT 04-SEP-2000 (first entry)
XX
DE A transcription factor designated CATFIIIA.
XX
KW Transcription factor; CATFIIIA; DNA-binding protein;
KW ribosomal RNA 5S gene; fungal infection.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT Misc-difference 193
FT /note= "Ser encoded by CTG"
FT Misc-difference 339
FT /note= "Ser encoded by CTG"
XX
PN WO200028037-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002739.
XX
PR 10-NOV-1998; 98FR-00014147.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Camier S, Sentenac A;
XX
DR WPI; 2000-376549/32.
DR N-PSDB; AAA15398.
XX
PT New nucleic acid encoding Candida albicans transcription factor, useful
PT e.g. in screening for antimycotic agents and for immunization.
XX
PS Claim 12; Page 35-36; 45pp; French.
XX
CC The present sequence represents a Candida albicans transcription factor,
CC designated CATFIIIA. The polypeptide is a DNA-binding protein, which is
CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as antimycotic agents, to raise an antibody response that is
CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in

```
CC   compositions for diagnosing and treating fungal infections, e.g. by
CC   detecting polymorphisms and mutations
XX
SQ   Sequence 412 AA;

  Query Match            100.0%;  Score 2229;  DB 3;  Length 412;
  Best Local Similarity  100.0%;  Pred. No. 1e-167;
  Matches  412;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180

Qy        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240

Qy        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360

Qy        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412


RESULT 2
ABP73656
ID   ABP73656 standard; protein; 412 AA.
XX
AC   ABP73656;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential protein SEQ ID NO 7493.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   N-PSDB; ABZ32206.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans protein used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office
XX
SQ   Sequence 412 AA;

  Query Match            99.8%;  Score 2225;  DB 5;  Length 412;
  Best Local Similarity  99.8%;  Pred. No. 2.2e-167;
  Matches  411;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVE 60

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180

Qy        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240

Qy        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360

Qy        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412


RESULT 3
ABJ25655
ID   ABJ25655 standard; protein; 564 AA.
XX
AC   ABJ25655;
XX
DT   16-APR-2003  (first entry)
XX
```

DE Aspergillus fumigatus essential gene protein #313.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 564 AA;

Query Match 20.0%; Score 445; DB 6; Length 564;
Best Local Similarity 30.3%; Pred. No. 2.3e-26;
Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

```
Qy          14 ISSSSSSRP---KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
               ||::|:  |   | : | ::|| ||:|||: |::|||:|:|:| :||| ::||| | | |
Db          71 ISTTSAKYPSELKTHRCPFDGCTKAFNRPARLQEHLRSHNNERIFKCTFEECDKTFLRAS 130

Qy          71 HLETHIVS-HSEKKPFHCS--VCGKGVNSRQHLKRHEITH--TKSFKCT-FENCQEAFYK 124
               ||  || | |:  : : |    |||   :   |:||   |     ::|| :  | | | |
Db         131 HLNHHIKSAHTGVRDYVCDRPGCGKSFVTGSRLRRHLAAHDGRDKYRCTEYPPCNETFRK 190

Qy         125 HQSLRHHILSVH--EKTLTC--------KQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH- 173
               |:|: ||:: |  :|   |         ::|   |      |  |: : |      : |
Db         191 HSTLQKHIMTAHLKQKPFQCPHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFSCTEC 249

Qy         174 ----PGCFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHM-LSHDDSTM--IK 226
                    |    | |:::|| ||:  ||  :|| |   |   : |  |: ::| | ::   |
Db         250 SQHAEGAEATFPTYALLQAHIRSVHPP-QCPNCALTCATSRELRRHLEVAHGDVSLEERK 308

Qy         227 IWTCDY--CDVGKFAKKNELVEHYNIFHDGNI-----PDDLLKETEVKKLENLLDQ---- 275
               |: |    ||   | ||  |  |    | |           ||    :|    |  |
Db         309 IFPCTVPGCD-RSFTKKGNLTVHIRTVHQGEKRFVCGETDLSSSKKVSGWNN--DNGCGK 365

Qy         276 --GSKLNNLHELETEKLKVEEDEEDEEDSL----DEKRSDVRSDSMSAQRSIKSFTASLE 329
                 ||||     : |  |  :  : :    |    | : |   |  :||        |:|
Db         366 RYGSKLALEEHIRTAHLGYQNAKAERRQRLGITRDRQHSTATSPGVSA-------LAALT 418

Qy         330 GSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHL--KWHDDNLQRIESFL 380
               |     :   :|: | |   :|   | |:|||  |:  | |    :  : ||
Db         419 GEGYAEE----TGRHIACLVESCPHRFHRDYDLWVHMSGKHHFSEEETRDLFL 467
```

RESULT 4
ABJ26255
ID ABJ26255 standard; protein; 564 AA.
XX
AC ABJ26255;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #913.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and

CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 564 AA;

Query Match 20.0%; Score 445; DB 6; Length 564;
Best Local Similarity 30.3%; Pred. No. 2.3e-26;
Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

Qy 14 ISSSSSSRP---KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
Db 71 ISTTSAKYPSELKTHRCPFDGCTKAFNRPARLQEHLRSHNNERIFKCTFEECDKTFLRAS 130

Qy 71 HLETHIVS-HSEKKPFHCS--VCGKGVNSRQHLKRHEITH--TKSFKCT-FENCQEAFYK 124
Db 131 HLNHHIKSAHTGVRDYVCDRPGCGKSFVTGSRLRRHLAAHDGRDKYRCTEYPPCNETFRK 190

Qy 125 HQSLRHHILSVH--EKTLTC--------KQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH- 173
Db 191 HSTLQKHIMTAHLKQKPFQCPHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFSCTEC 249

Qy 174 ----PGCFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHM-LSHDDSTM--IK 226
Db 250 SQHAEGAEATFPTYALLQAHIRSVHPP-QCPNCALTCATSRELRRHLEVAHGDVSLEERK 308

Qy 227 IWTCDY--CDVGKFAKKNELVEHYNIFHDGNI-----PDDLLKETEVKKLENLLDQ---- 275
Db 309 IFPCTVPGCD-RSFTKKGNLTVHIRTVHQGEKRFVCGETDLSSSKKVSGWNN--DNGCGK 365

Qy 276 --GSKLNNLHELETEKLKVEEDEEDEEDSL----DEKRSDVRSDSMSAQRSIKSFTASLE 329
Db 366 RYGSKLALEEHIRTAHLGYQNAKAERRQRLGITRDRQHSTATSPGVSA-------LAALT 418

Qy 330 GSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHL--KWHDDNLQRIESFL 380
Db 419 GEGYAEE----TGRHIACLVESCPHRFHRDYDLWVHMSGKHHFSEEETRDLFL 467

RESULT 5
ABG00399
ID ABG00399 standard; protein; 1212 AA.
XX
AC ABG00399;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #390.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS64586.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 30758; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1212 AA;

Query Match 18.7%; Score 416.5; DB 4; Length 1212;
Best Local Similarity 32.6%; Pred. No. 1.1e-23;
Matches 116; Conservative 45; Mismatches 148; Indels 47; Gaps 15;

Qy 20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
Db 861 TREKTYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 916

Qy 80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
Db 917 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 974

Qy 137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
Db 975 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1030

Qy 197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253
Db 1031 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1085

Qy 254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSD 313
Db 1086 GE------KPYKCEKCGKAFNQSSILTNHKKIHTITPKIH----------------TREK 1123

Qy 314 SMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
Db 1124 PYKYKECGKSFNRSSTFTK--HKVIHTGVKLYKC--EECGKSFFWSSALTRHKKIH 1175

RESULT 6
ADN99509
ID ADN99509 standard; protein; 664 AA.
XX
AC ADN99509;
XX
DT 29-JUL-2004 (first entry)

```
XX
DE   Novel human protein sequence #325.
XX
KW   anti-inflammatory; dermatological; neuroprotective; immunomodulator;
KW   antibacterial; virucide; antipsoriatic; cytostatic; gene therapy;
KW   vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
KW   early aging; hormonal imbalance; ischemic heart disease;
KW   ulcerative colitis.
XX
OS   Homo sapiens.
XX
PN   WO2004038003-A2.
XX
PD   06-MAY-2004.
XX
PF   24-OCT-2003; 2003WO-US033947.
XX
PR   25-OCT-2002; 2002US-0421061P.
PR   25-OCT-2002; 2002US-0421080P.
PR   25-OCT-2002; 2002US-0421552P.
PR   25-OCT-2002; 2002US-0421614P.
PR   30-OCT-2002; 2002US-0422177P.
PR   30-OCT-2002; 2002US-0422178P.
PR   15-NOV-2002; 2002US-0426355P.
PR   15-NOV-2002; 2002US-0426384P.
PR   15-NOV-2002; 2002US-0426394P.
PR   15-NOV-2002; 2002US-0426430P.
PR   15-NOV-2002; 2002US-0426916P.
PR   27-NOV-2002; 2002US-0429224P.
PR   27-NOV-2002; 2002US-0429275P.
PR   27-NOV-2002; 2002US-0429302P.
PR   27-NOV-2002; 2002US-0429326P.
PR   27-NOV-2002; 2002US-0429651P.
PR   04-DEC-2002; 2002US-0430645P.
PR   04-DEC-2002; 2002US-0430651P.
PR   04-DEC-2002; 2002US-0430657P.
PR   04-DEC-2002; 2002US-0430663P.
PR   04-DEC-2002; 2002US-0430668P.
PR   04-DEC-2002; 2002US-0430684P.
PR   05-DEC-2002; 2002US-0430937P.
PR   05-DEC-2002; 2002US-0430965P.
PR   05-DEC-2002; 2002US-0431458P.
PR   12-DEC-2002; 2002US-0433251P.
PR   12-DEC-2002; 2002US-0433500P.
PR   13-DEC-2002; 2002US-0433316P.
PR   13-DEC-2002; 2002US-0433318P.
PR   23-DEC-2002; 2002US-0436236P.
PR   03-JAN-2003; 2003US-0437914P.
PR   17-JAN-2003; 2003US-0440820P.
PR   17-JAN-2003; 2003US-0440821P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476621P.
PR   09-JUN-2003; 2003US-0476632P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485217P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   08-JUL-2003; 2003US-0485359P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Kothakota S,  Lin H,  Linnemann T,  Pierce K,  Wang Y;
PI   Wong JGP,  Wu G,  Zhang H,  Zeng C;
XX
DR   WPI; 2004-365511/34.
DR   N-PSDB; ADN98725.
XX
PT   New nucleic acid molecules, useful in preparing a composition for
PT   treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT   disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT   ulcerative colitis.
XX
PS   Claim 14; SEQ ID NO 1109; 532pp; English.
XX
CC   The invention relates to a nucleic acid molecule comprising a
CC   polynucleotide sequence or its complement that encodes a polypeptide. The
CC   nucleic acid is useful in preparing a composition for treating or
CC   preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC   cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC   heart disease or ulcerative colitis. This sequence corresponds to a
CC   protein of the invention.
XX
SQ   Sequence 664 AA;

  Query Match             18.3%;  Score 407;  DB 8;  Length 664;
  Best Local Similarity   32.6%;  Pred. No. 2.9e-23;
  Matches  115;  Conservative   46;  Mismatches  138;  Indels   54;  Gaps   16;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  : ||| ::| | ||:| | |: ::|||| |  |||||   |||  ||| |: :
Db        298 KPYQC--KECDKVFSRKSYLERHRRIHTGEKPYKCKV--CDKAFRHDSHLAQHIVIHTRE 353

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHEKT 139
              ||: |: |||        |  |:  ||  | :||    | : | :  :| ||| |   ||
Db        354 KPYKCNECGKTFGENSALLVHKTIHTGEKPYKC--NECGKVFNQQSNLARHHRLHTGEKP 411

Qy        140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
             ||:|:|||:| | | :|:  | |  | |:|    | | |:  | |  |  |
Db        412 YKCKECDKVFSRKSHLERHRRIHTGEKP-YKC--KVCDKAFRRDSHLAQHTVIHTGEKPY 468

Qy        198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNI 256
             || :|||  |    |  | : |   |  | : |: |  || |  |: |  |: : | |
Db        469 KCNECGKTFVQNSSLVMHKVIH---TGEKRYKCNEC--GKSFNHKSSLAYHHRL-HTGEK 522

Qy        257 PDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMS 316
             |    |  |  |:     | :  : ||  | :  | || ::
Db        523 P---YKCNECGKVFRTQSQLACHHRLHTGE-KPYKCEECDK------------------ 559

Qy        317 AQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                      :::   : : :    |   |    ||: | |:  | :| : |
Db        560 --------VFNIKSHLEIHRRVHTGEKPYKC--RVCDKAFGRDSYLAQHQRVH 602


RESULT 7
AAU30831
ID   AAU30831 standard; protein; 1230 AA.
XX
AC   AAU30831;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #1322.
XX
```

KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 355; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 1230 AA;

Query Match 18.1%; Score 403; DB 4; Length 1230;
Best Local Similarity 32.9%; Pred. No. 1.4e-22;
Matches 118; Conservative 45; Mismatches 144; Indels 52; Gaps 17;

```
Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
Db        891 TREKPYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 946

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
Db        947 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1004

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
Db       1005 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1060

Qy        197 LKCPKCGKGCVGK------KGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNI 250
Db       1061 EKPYKC-EGC-GKAFSQSFNTLTGHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKI 1113

Qy        251 FHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDV 310
Db       1114 IHTGE------KPYKCEKCCKAFNQSSILTNHKKIHTITPKIH----------------T 1151

Qy        311 RSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
Db       1152 REKPYKYKECGKSFNRSSTFTK--HKVIHTGVKLYKC--EECGKSFFWSSALTRHKKIH 1206
```

RESULT 8
ADM05916
ID ADM05916 standard; protein; 404 AA.
XX
AC ADM05916;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human protein of the invention SEQ ID NO:4601.
XX
KW human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR N-PSDB; ADM03473.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 4601; 305pp; English.
XX
CC The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC protein sequence of the invention.
XX
SQ Sequence 404 AA;

Query Match 17.9%; Score 400; DB 7; Length 404;
Best Local Similarity 33.9%; Pred. No. 5.2e-23;
Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;

```
Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db         39 KPYKC--EECGKAFNRSSILTKHKIIHTGEKPYKC--EECGKGFSSVSTLNTHKAIHAEE 94

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSVH--EK 138
Db         95 KPYKCEECGKASNSSSKLMEHKRIHTGEKPYKC--EECGKAFSWSSSLTEH-KRIHAGEK 151

Qy        139 TLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPK 196
Db        152 PYKCEECGKAFNRSSILTKHKIIHTGEKP-YKCE--GCGKAFSKVSTLNTHKAIHAEEKP 208

Qy        197 LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGN 255
Db        209 YKCEECGKASNSSSKLMEHKRIH---TGEKPYKCEEC--GKAFSWSSSLTEHKRI-HAGE 262

Qy        256 IPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEE-DEEDEEDSLDEKRSDVRSDS 314
```

```
Db        263 KP---YKCEECGKAFTWSSSFTKHKRIHAAE-KPYKCEECGKGFSTFSILTKHKIIHTGE 318

Qy        315 --MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369

Db        319 KRYKCEECGKAF--SWSSILTEHKIIHTGEKPYKC--EECGKAFSRSSSLTRHKRIH 371


RESULT 9
ADN06034
ID   ADN06034 standard; protein; 803 AA.
XX
AC   ADN06034;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Antipsoriatic protein sequence #1173.
XX
KW   antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO2004028479-A2.
XX
PD   08-APR-2004.
XX
PF   25-SEP-2003; 2003WO-US030907.
XX
PR   25-SEP-2002; 2002US-0414006P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Bodary S,  Clark H,  Jackman J,  Schoenfeld J,  Williams PM,  Wood WI;
PI   Wu TD;
XX
DR   WPI; 2004-305105/28.
DR   N-PSDB; ADN06033.
XX
PT   New PRO nucleic acid or polypeptide, useful for preparing a
PT   pharmaceutical composition for diagnosing or treating psoriasis in a
PT   mammal.
XX
PS   Claim 9; SEQ ID NO 2429; 3069pp; English.
XX
CC   The invention relates to novel polynucleotide and polypeptides for
CC   treating psoriasis or a sequence having at least 80% identity to the
CC   above sequences. The nucleic acid is useful for preparing a composition
CC   for diagnosing or treating psoriasis in a mammal. This sequence
CC   corresponds to one of the polypeptides of the invention.
XX
SQ   Sequence 803 AA;

  Query Match             17.9%;  Score 400;  DB 8;  Length 803;
  Best Local Similarity   31.9%;  Pred. No. 1.3e-22;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59

Db        367 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 422

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117

Db        423 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 480

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176

Db        481 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 537

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234

Db        538 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 593

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293

Db        594 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 647

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350

Db        648 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 704

Qy        351 NCDRMFSREYDLRRHLKWH 369

Db        705 -CGKAFNRPSNLIEHKKIH 722


RESULT 10
AAM38689
ID   AAM38689 standard; protein; 809 AA.
XX
AC   AAM38689;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1834.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J,  Zhao QA;
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI57845.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Example 3; SEQ ID NO 1834; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
```

```
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 809 AA;

  Query Match             17.9%;  Score 400;  DB 4;  Length 809;
  Best Local Similarity   31.9%;  Pred. No. 1.3e-22;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59
              :|  |  | || ::      :  | | |  | | ||:   | | :|  ||: ::||||
Db        373 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 428

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSPKCTFEN 117
              ::| |||   | |  |   |: :||: | ||||  |   :|  |:  ||  | :||  |
Db        429 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 486

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176
              | :|| :  :| :|  : : :|   |::| | |   ||| :||: | |  | |:|:   |
Db        487 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 543

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234
              | |  :|:|  |  |        || :|||       |::|   |   |  | : |: |
Db        544 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 599

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293
              || | : : |  | | | |  |     | |  |  |  |     :|   :|  | :  | |
Db        600 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 653

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350
              |  :  :  |   |  : :      :   |:|   |  : |  |:|    |   | |
Db        654 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 710

Qy        351 NCDRMFSREYDLRRHLKWH 369
              | : |:|  :|  | | |
Db        711 -CGKAFNRPSNLIEHKKIH 728


RESULT 11
ADN99691
ID   ADN99691 standard; protein; 444 AA.
XX
AC   ADN99691;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Novel human protein sequence #507.
XX
KW   anti-inflammatory; dermatological; neuroprotective; immunomodulator;
KW   antibacterial; virucide; antipsoriatic; cytostatic; gene therapy;
KW   vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
KW   early aging; hormonal imbalance; ischemic heart disease;
KW   ulcerative colitis.
XX
OS   Homo sapiens.
XX
PN   WO2004038003-A2.
XX
PD   06-MAY-2004.
XX
PF   24-OCT-2003; 2003WO-US033947.
XX
PR   25-OCT-2002; 2002US-0421061P.
PR   25-OCT-2002; 2002US-0421080P.
PR   25-OCT-2002; 2002US-0421552P.
PR   25-OCT-2002; 2002US-0421614P.
PR   30-OCT-2002; 2002US-0422177P.
PR   30-OCT-2002; 2002US-0422178P.
PR   15-NOV-2002; 2002US-0426355P.
PR   15-NOV-2002; 2002US-0426384P.
PR   15-NOV-2002; 2002US-0426394P.
PR   15-NOV-2002; 2002US-0426430P.
PR   15-NOV-2002; 2002US-0426916P.
PR   27-NOV-2002; 2002US-0429224P.
PR   27-NOV-2002; 2002US-0429275P.
PR   27-NOV-2002; 2002US-0429302P.
PR   27-NOV-2002; 2002US-0429326P.
PR   27-NOV-2002; 2002US-0429651P.
PR   04-DEC-2002; 2002US-0430645P.
PR   04-DEC-2002; 2002US-0430651P.
PR   04-DEC-2002; 2002US-0430657P.
PR   04-DEC-2002; 2002US-0430663P.
PR   04-DEC-2002; 2002US-0430668P.
PR   04-DEC-2002; 2002US-0430684P.
PR   05-DEC-2002; 2002US-0430937P.
PR   05-DEC-2002; 2002US-0430965P.
PR   05-DEC-2002; 2002US-0431458P.
PR   12-DEC-2002; 2002US-0433251P.
PR   12-DEC-2002; 2002US-0433500P.
PR   13-DEC-2002; 2002US-0433316P.
PR   13-DEC-2002; 2002US-0433318P.
PR   23-DEC-2002; 2002US-0436236P.
PR   03-JAN-2003; 2003US-0437914P.
PR   17-JAN-2003; 2003US-0440820P.
PR   17-JAN-2003; 2003US-0440821P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476621P.
PR   09-JUN-2003; 2003US-0476632P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485217P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   08-JUL-2003; 2003US-0485359P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Kothakota S,  Lin H,  Linnemann T,  Pierce K,  Wang Y;
PI   Wong JGP,  Wu G,  Zhang H,  Zeng C;
XX
DR   WPI; 2004-365511/34.
DR   N-PSDB; ADN98907.
XX
PT   New nucleic acid molecules, useful in preparing a composition for
PT   treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT   disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT   ulcerative colitis.
XX
PS   Claim 14; SEQ ID NO 1291; 532pp; English.
XX
CC   The invention relates to a nucleic acid molecule comprising a
CC   polynucleotide sequence or its complement that encodes a polypeptide. The
CC   nucleic acid is useful in preparing a composition for treating or
```

```
CC   preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC   cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC   heart disease or ulcerative colitis. This sequence corresponds to a
CC   protein of the invention.
XX
SQ   Sequence 444 AA;

  Query Match             17.9%;  Score 399;  DB 8;  Length 444;
  Best Local Similarity   30.8%;  Pred. No. 7.1e-23;
  Matches  120;  Conservative   47;  Mismatches  179;  Indels   44;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
Db         77 SGEKPYKC--KECGKAYNEASNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 132

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
Db        133 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 190

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
Db        191 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSRLSHLTTHKRIHSGE 247

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
Db        248 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 302

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEE-DEEDEEDSLDEKRSDVRSD 313
Db        303 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEECGKAFNQSSTLSKHKVIHTG 358

Qy        314 S--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWHD- 370
Db        359 EKPYKCEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSILNRHKMIHTG 414

Qy        371 ------------DNLQRIESFLNSIEKEE 387
Db        415 EKLYKPESCNNACDNIAKISKYKRNCAGEK 444


RESULT 12
ADN99690
ID   ADN99690 standard; protein; 444 AA.
XX
AC   ADN99690;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Novel human protein sequence #506.
XX
KW   anti-inflammatory; dermatological; neuroprotective; immunomodulator;
KW   antibacterial; virucide; antipsoriatic; cytostatic; gene therapy;
KW   vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
KW   early aging; hormonal imbalance; ischemic heart disease;
KW   ulcerative colitis.
XX
OS   Homo sapiens.
XX
PN   WO2004038003-A2.
XX
PD   06-MAY-2004.
XX
PF   24-OCT-2003; 2003WO-US033947.
XX
PR   25-OCT-2002; 2002US-0421061P.
PR   25-OCT-2002; 2002US-0421080P.
PR   25-OCT-2002; 2002US-0421552P.
PR   25-OCT-2002; 2002US-0421614P.
PR   30-OCT-2002; 2002US-0422177P.
PR   30-OCT-2002; 2002US-0422178P.
PR   15-NOV-2002; 2002US-0426355P.
PR   15-NOV-2002; 2002US-0426384P.
PR   15-NOV-2002; 2002US-0426394P.
PR   15-NOV-2002; 2002US-0426430P.
PR   15-NOV-2002; 2002US-0426916P.
PR   27-NOV-2002; 2002US-0429224P.
PR   27-NOV-2002; 2002US-0429275P.
PR   27-NOV-2002; 2002US-0429302P.
PR   27-NOV-2002; 2002US-0429326P.
PR   27-NOV-2002; 2002US-0429651P.
PR   04-DEC-2002; 2002US-0430645P.
PR   04-DEC-2002; 2002US-0430651P.
PR   04-DEC-2002; 2002US-0430657P.
PR   04-DEC-2002; 2002US-0430663P.
PR   04-DEC-2002; 2002US-0430668P.
PR   04-DEC-2002; 2002US-0430684P.
PR   05-DEC-2002; 2002US-0430937P.
PR   05-DEC-2002; 2002US-0430965P.
PR   05-DEC-2002; 2002US-0431458P.
PR   12-DEC-2002; 2002US-0433251P.
PR   12-DEC-2002; 2002US-0433500P.
PR   13-DEC-2002; 2002US-0433316P.
PR   13-DEC-2002; 2002US-0433318P.
PR   23-DEC-2002; 2002US-0436236P.
PR   03-JAN-2003; 2003US-0437914P.
PR   17-JAN-2003; 2003US-0440820P.
PR   17-JAN-2003; 2003US-0440821P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476621P.
PR   09-JUN-2003; 2003US-0476632P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485217P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   08-JUL-2003; 2003US-0485359P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Kothakota S,  Lin H,  Linnemann T,  Pierce K,  Wang Y;
PI   Wong JGP,  Wu G,  Zhang H,  Zeng C;
XX
DR   WPI; 2004-365511/34.
DR   N-PSDB; ADN98906.
XX
PT   New nucleic acid molecules, useful in preparing a composition for
PT   treating or preventing e.g. inflammatory, CNS, bacterial or viral
PT   disorders, cancer, psoriasis, diabetes, ischemic heart disease or
PT   ulcerative colitis.
XX
PS   Claim 14; SEQ ID NO 1290; 532pp; English.
XX
CC   The invention relates to a nucleic acid molecule comprising a
CC   polynucleotide sequence or its complement that encodes a polypeptide. The
```

```
CC   nucleic acid is useful in preparing a composition for treating or
CC   preventing inflammatory, CNS, immune, bacterial or viral disorder,
CC   cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
CC   heart disease or ulcerative colitis. This sequence corresponds to a
CC   protein of the invention.
XX
SQ   Sequence 444 AA;

  Query Match            17.9%;  Score 398;  DB 8;  Length 444;
  Best Local Similarity  30.2%;  Pred. No. 8.5e-23;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              |  | | |  : | ||||  | |  | | |:  :|||| ::| ||| | ||| || : |
Db         77 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 132

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : |||: |  |||  |   :|  |:  ||  | :||  | |  || :  :|  | |:
Db        133 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 190

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
              ||   |::| | |:: | |  ||:  | |   |:|:   | | |   | | |  |
Db        191 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 247

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
                || :|||      |::|   |      | : |: |   |:: : | |  | | |
Db        248 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 302

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
                |    |  | |  ||  | :    :|  | :  | ||            ::  :| :
Db        303 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 348

Qy        315 MSAQRSI------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
              :|  : |             |:|  |     :  |:|    |   |    | : |:
Db        349 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 404

Qy        362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387
              | ||  |               ||: :|  :  :   |:
Db        405 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 444


RESULT 13
ABO60215
ID   ABO60215 standard; protein; 487 AA.
XX
AC   ABO60215;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon protein #6449.
XX
KW   Human; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 45; SEQ ID NO 33849; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe protein of the invention. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 487 AA;

  Query Match            17.9%;  Score 398;  DB 8;  Length 487;
  Best Local Similarity  30.2%;  Pred. No. 9.7e-23;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              |  | | |  : | ||||  | |  | | |:  :|||| ::| ||| | ||| || : |
Db        120 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 175

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : |||: |  |||  |   :|  |:  ||  | :||  | |  || :  :|  | |:
Db        176 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 233

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
              ||   |::| | |:: | |  ||:  | |   |:|:   | | |   | | |  |
Db        234 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 290

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
                || :|||      |::|   |      | : |: |   |:: : | |  | | |
Db        291 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 345

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
                |    |  | |  ||  | :    :|  | :  | ||            ::  :| :
Db        346 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 391

Qy        315 MSAQRSI------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
              :|  : |             |:|  |     :  |:|    |   |    | : |:
Db        392 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 447

Qy        362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387
```

```
                    | ||   |                ||: :|  :  :   |:
Db        448 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 487
```

RESULT 14
ABU96690
ID   ABU96690 standard; protein; 563 AA.
XX
AC   ABU96690;
XX
DT   25-JUL-2003 (first entry)
XX
DE   Human nucleic acid-associated protein (NAAP) #19.
XX
KW   Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
KW   anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW   antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW   cell proliferative disorder; cancer; atherosclerosis;
KW   neurological disorder; epilepsy; Huntington's disease; stroke;
KW   immune disorder; inflammatory disorder; AIDS; allergy;
KW   developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW   protein-protein interaction; drug-target interaction;
KW   gene expression profile.
XX
OS   Homo sapiens.
XX
PN   WO2003023003-A2.
XX
PD   20-MAR-2003.
XX
PF   05-SEP-2002; 2002WO-US028540.
XX
PR   07-SEP-2001; 2001US-0317792P.
PR   07-SEP-2001; 2001US-0317912P.
PR   14-SEP-2001; 2001US-0322270P.
PR   21-SEP-2001; 2001US-0324040P.
PR   28-SEP-2001; 2001US-0326732P.
PR   19-OCT-2001; 2001US-0346716P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ;
PI   Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BM;
PI   Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI   Hafalia AJA, Sanjanwala B, Marquis JP, Gorvad AE, Lee SY, Ison CH;
PI   Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebarjadian Y, Shah P;
PI   Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE;
PI   Burford N, Ramkumar J;
XX
DR   WPI; 2003-313243/30.
DR   N-PSDB; ACA98938.
XX
PT   New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT   infections.
XX
PS   Claim 1; Page 255-257; 345pp; English.
XX
CC   The invention describes a novel human isolated nucleic acid-associated
CC   polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC   diagnosing, treating and preventing diseases or conditions associated
CC   with the decreased expression or overexpression of NAAP, such as cell
CC   proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC   epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC   allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC   disorders, or infections. These are also useful in assessing the effects
CC   of exogenous compounds on the expression of nucleic acid and amino acid
CC   sequences of NAAP. The NAAP or its fragments are useful in screening
CC   compounds for effectiveness as agonist or antagonist of the polypeptides,
CC   or in altering the expression of the target polynucleotide and compounds
CC   that specifically bind to or modulate the activity of the polypeptide.
CC   The microarray is useful in monitoring or measuring protein-protein
CC   interactions, drug-target interactions, and gene expression profiles.
CC   This is the amino acid sequence of a novel human nucleic acid-associated
CC   protein (NAAP)
XX
SQ   Sequence 563 AA;

```
  Query Match             17.9%;  Score 398;  DB 6;  Length 563;
  Best Local Similarity   30.2%;  Pred. No. 1.2e-22;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              |  | | |  : | ||||  | |  | | |:  :|||| ::| ||| | ||| || : |
Db        196 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 251

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : |||: |  |||  |   :|  |:  ||  | :||  | |  || :  :|  | |:
Db        252 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 309

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
              ||   |::| | |:: | |  ||:  | |   |:|:   | | |   | |  |  |
Db        310 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 366

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
                 || :|||       |::|   |       | : |:|    |:: : |  |  | | |
Db        367 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 421

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
                |   |  |  |  ||  | :    :|  | :  | ||            ::  :| :
Db        422 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 467

Qy        315 MSAQRSI------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
              :|  : |            |:|  |     :  |:|    |   |    | : |:
Db        468 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 523

Qy        362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387
              | ||   |              ||: :|  :  :   |:
Db        524 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 563
```

RESULT 15
ADI60246
ID   ADI60246 standard; protein; 563 AA.
XX
AC   ADI60246;
XX
DT   15-APR-2004 (first entry)
XX
DE   Secreted polypeptide #130.
XX
KW   osteopathic; vulnerary; cytostatic; gene therapy; diagnosis; forensics;
KW   gene mapping; mutation identification; biodiversity; chromosome marker;
KW   immune response; myeloid cell disorder; lymphoid cell disorder;
KW   bone cartilage; tendon; ligament; nerve tissue growth; wound healing;
KW   burns; incision; ulcer; cancer.
XX
OS   Homo sapiens.
XX
PN   WO2003025142-A2.
XX
PD   27-MAR-2003.
XX
PF   18-SEP-2002; 2002WO-US029636.
XX
PR   18-SEP-2001; 2001US-0323349P.
PR   16-SEP-2002; 2002US-00323349.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI   Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT;
XX

DR WPI; 2003-354601/33.
DR N-PSDB; ADI60591.
XX
PT New polynucleotides and secreted proteins, useful for treating myeloid or
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
PT tissue growth or regeneration, in wound healing, and in tissue repair and
PT replacement.
XX
PS Claim 20; SEQ ID NO 281; 243pp; English.
XX
CC The invention relates to novel isolated polynucleotides or a sequence
CC encoding a polypeptide with biological activity, where the polynucleotide
CC hybridizes to the polynucleotide under stringent hybridization conditions
CC or has greater than 99% sequence identity with the polynucleotide. The
CC polynucleotides and polypeptides are useful in diagnostics, forensics,
CC gene mapping, identification of mutations responsible for genetic
CC disorders and other traits, to assess biodiversity, as nutritional
CC sources or supplements. The polynucleotides may also be used as molecular
CC weight markers, chromosome markers or map related gene positions, or as
CC an antigen to raise anti-DNA antibodies or elicit immune response. The
CC polypeptides are useful for raising antibodies, as markers for tissues in
CC which the corresponding polypeptide is expressed, for re-engineering
CC damaged or diseased tissues, for treating myeloid or lymphoid cell
CC disorders, in bone cartilage, tendon, ligament and/or nerve tissue growth
CC or regeneration, in wound healing, in tissue repair and replacement, in
CC healing of burns, incisions and ulcers, and in treating cancer. This
CC sequence corresponds to a protein sequence of the invention.
XX
SQ Sequence 563 AA;

Query Match 17.9%; Score 398; DB 7; Length 563;
Best Local Similarity 30.2%; Pred. No. 1.2e-22;
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

```
Qy          20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
               |  | | |  : | ||||  | |  | | |:  :||||  ::| ||| | ||| || : |
Db         196 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 251

Qy          80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSPKCTFENCQEAFYKHQSL-RHHILSVH 136
               : |||: |  |||  |   :|  |:  ||  | :||  | |  || :  :|  | |:
Db         252 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 309

Qy         137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
               ||   |::| | |:: | |  ||:  | |   |:|:   | | |   | |  |  |
Db         310 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 366

Qy         195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
                  || :|||       |::|   |       | : |: |    |:: : |  |  | | |
Db         367 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 421

Qy         255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
                  |    |  |  |  ||  | :    :|  | :  | ||           ::  :| :
Db         422 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 467

Qy         315 MSAQRSI------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
               :|  : |            |:|  |     :  |:|    |   |    | : |:
Db         468 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 523

Qy         362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387
               | ||   |               ||: :|  :  :   |:
Db         524 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 563
```

Search completed: February 8, 2005, 15:18:59
Job time : 79 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:14:56 ; Search time 42 Seconds
(without alignments)
732.272 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1036	46.5	208	4	US-09-248-796A-19287	Sequence 19287, A
2	671	30.1	153	4	US-09-248-796A-19289	Sequence 19289, A
3	400	17.9	803	4	US-09-538-092-1026	Sequence 1026, Ap
4	399.5	17.9	656	4	US-09-949-016-10075	Sequence 10075, A
5	395.5	17.7	423	1	US-08-523-376-3	Sequence 3, Appli
6	393.5	17.7	1191	4	US-09-949-016-6356	Sequence 6356, Ap
7	358	16.1	457	4	US-09-389-956-68	Sequence 68, Appl
8	358	16.1	647	4	US-09-389-956-6	Sequence 6, Appli
9	351	15.7	487	4	US-09-949-016-10055	Sequence 10055, A
10	350	15.7	685	4	US-09-949-016-8627	Sequence 8627, Ap
11	344	15.4	639	4	US-09-949-016-9434	Sequence 9434, Ap
12	343	15.4	746	4	US-09-949-016-11494	Sequence 11494, A
13	343	15.4	769	4	US-09-949-016-8149	Sequence 8149, Ap
14	343	15.4	810	4	US-09-538-092-1275	Sequence 1275, Ap
15	337	15.1	469	4	US-09-538-092-1195	Sequence 1195, Ap
16	333	14.9	642	4	US-09-949-016-6986	Sequence 6986, Ap
17	332	14.9	648	4	US-09-949-016-10374	Sequence 10374, A
18	329	14.8	589	4	US-09-538-092-1194	Sequence 1194, Ap
19	329	14.8	728	2	US-08-475-844-5	Sequence 5, Appli
20	329	14.8	728	5	PCT-US95-08429-5	Sequence 5, Appli
21	322	14.4	727	2	US-08-475-844-9	Sequence 9, Appli
22	322	14.4	727	5	PCT-US95-08429-9	Sequence 9, Appli
23	322	14.4	736	4	US-09-949-016-9241	Sequence 9241, Ap
24	321	14.4	654	4	US-09-538-092-1193	Sequence 1193, Ap
25	321	14.4	654	4	US-09-949-016-6357	Sequence 6357, Ap
26	321	14.4	663	4	US-09-949-016-7484	Sequence 7484, Ap
27	319.5	14.3	711	2	US-08-820-170A-10	Sequence 10, Appl
28	319.5	14.3	711	3	US-09-055-699-10	Sequence 10, Appl
29	319.5	14.3	711	3	US-09-273-565-10	Sequence 10, Appl
30	319.5	14.3	711	3	US-09-565-538-10	Sequence 10, Appl
31	319.5	14.3	711	3	US-09-661-468-10	Sequence 10, Appl
32	319.5	14.3	711	4	US-09-976-165-10	Sequence 10, Appl
33	319.5	14.3	711	4	US-09-828-648-2	Sequence 2, Appli
34	317	14.2	458	4	US-09-538-092-1356	Sequence 1356, Ap
35	316.5	14.2	321	4	US-09-538-092-1304	Sequence 1304, Ap
36	316.5	14.2	321	4	US-09-949-016-6364	Sequence 6364, Ap
37	316	14.2	488	2	US-08-933-750C-17	Sequence 17, Appl
38	316	14.2	488	3	US-09-234-613-17	Sequence 17, Appl
39	316	14.2	518	4	US-09-881-578A-4	Sequence 4, Appli
40	315.5	14.2	334	4	US-09-949-016-8516	Sequence 8516, Ap
41	315	14.1	803	3	US-09-063-035-2	Sequence 2, Appli
42	314.5	14.1	532	4	US-09-389-956-80	Sequence 80, Appl
43	313	14.0	506	4	US-09-538-092-1177	Sequence 1177, Ap
44	313	14.0	717	4	US-09-881-578A-2	Sequence 2, Appli
45	312.5	14.0	543	4	US-09-362-123A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-248-796A-19287
; Sequence 19287, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19287
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19287

Query Match 46.5%; Score 1036; DB 4; Length 208;
Best Local Similarity 96.2%; Pred. No. 3.8e-90;
Matches 202; Conservative 2; Mismatches 2; Indels 4; Gaps 2;

```
Qy        204 KG-CVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLK 262
              || |:|||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db          2 KGLCLGKKGLSSHMLSHDDFTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLK 61

Qy        263 ETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQRSIK 322
              |||||||||||||||||||||||||||||   |||||||||:|||||||||||||||||
Db         62 ETEVKKLENLLDQGSKLNNLHELETEKLKV---EEDEEDSLDERRSDVRSDSMSAQRSIK 118

Qy        323 SFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWHDDNLQRIESFLNS 382
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        119 SFTASLEGSKSVSKLILNSGKKINCPKNNCDRMFSREYDLRRHLKWHDDNLQRIESFLNS 178

Qy        383 IEKEETPEGEPLVKKARMDLLPNETSVISR 412
              |||||||||||||||||||||||||||||||
Db        179 IEKEETPEGEPLVKKARMDLLPNETSVISR 208
```

RESULT 2
US-09-248-796A-19289
; Sequence 19289, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA

```
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19289
;   LENGTH: 153
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-19289

  Query Match             30.1%;  Score 671;  DB 4;  Length 153;
  Best Local Similarity   98.4%;  Pred. No. 9.9e-56;
  Matches  122;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         18 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVE 77

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 137

Qy        121 AFYK 124
              ||||
Db        138 TFYK 141


RESULT 3
US-09-538-092-1026
; Sequence 1026, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1026
;   LENGTH: 803
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P28160
US-09-538-092-1026

  Query Match             17.9%;  Score 400;  DB 4;  Length 803;
  Best Local Similarity   31.9%;  Pred. No. 5.4e-29;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59
              :|  |  | || ::      :  | | |  | | ||:   | | :|  ||: ::||||
Db        367 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 422

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117
              ::| |||   | |  |   |: :||: | ||||  |   :|  |:  ||  | :||  |
Db        423 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 480

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176
              | :|| :  :| :|  : : :|   |::| | |   ||| :||: | |  | |:|:   |
Db        481 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 537

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234
              | |  :|:|  |  |        || :|||       |::|  |   |  | : |: |
Db        538 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 593

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293
              || | : |  |  | | |  |    |  |  |  |     :|   :|  | :  | |
Db        594 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 647

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350
              |  :  :  |  |   : :      :  |:|  |  : |  |:|    |  | |
Db        648 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 704

Qy        351 NCDRMFSREYDLRRHLKWH 369
              | : |:|  :|  | | |
Db        705 -CGKAFNRPSNLIEHKKIH 722


RESULT 4
US-09-949-016-10075
; Sequence 10075, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10075
;   LENGTH: 656
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10075

  Query Match             17.9%;  Score 399.5;  DB 4;  Length 656;
  Best Local Similarity   29.7%;  Pred. No. 4.5e-29;
  Matches  116;  Conservative   56;  Mismatches  158;  Indels   61;  Gaps   19;

Qy         16 SSSSSRPKK-------YICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFR 68
              ||: :| |:        | |  | | ||::  | | :| | |: ::||||  ::| ||| |
Db         57 SSTLTRHKRIHTGEKPYKC--EECGKAFSHSSTLAKHKRIHTGEKPYKC--EECGKAFSR 112

Qy         69 KSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQ 126
              | |  |   |: :||: |  |||  ::   |  |:||||  | :||  : | : | :
Db        113 SSTLAKHKRIHTGEKPYKCKECGKAFSNSSTLANHKITHTEEKPYKC--KECDKTFKRLS 170

Qy        127 SL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSV 185
              :| :| |:   ||   |::| | | | | | ||  | | |  | |:|:   | | |   |
Db        171 TLTKHKIIHAGEKLYKCEECGKAFNRSSNLTIHKFIHTGEKP-YKCEE--CGKAFNWSSS 227

Qy        186 LQFHIKQSHPK---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKK 241
              |  | |: | :    || :|||  :    |: |   |    |  | : |: |  || |::
Db        228 LTKH-KRIHTREKPFKCKECGKAFIWSSTLTRHKRIH---TGEKPYKCEEC--GKAFSRS 281

Qy        242 NELVEHYNIFHDGNIP------DDLLKETEVKKLENLLDQGSKL----------NNLHEL 285
              : | :|  | | |  |           | :       ::  | ||           |    |
Db        282 STLTKHKTI-HTGEKPYKCKECGKAFKHSSALAKHKIIHAGEKLYKCEECGKAFNQSSNL 340

Qy        286 ETEKLKVEEDEEDEEDSLDE----------KRSDVRSDSMSAQRSIKSFTASLEGSKSV 334
              | |:   :::  : :  |:          ||   |  :   :   |:|  |     :
Db        341 TTHKIIHTKEKPSKSEECDKAFIWSSTLTEHKRIHTREKTYKCEECGKAF--SQPSHLTT 398

Qy        335 SKLISNSGKKINCPKNNCDRMFSREYDLRRH 365
```

```
                | :    |   |    | : ||:   | |
Db        399 HKRMHTGEKPYKC--EECGKAFSQSSTLTTH 427


RESULT 5
US-08-523-376-3
; Sequence 3, Application US/08523376
; Patent No. 5808030
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Satoshi, TAKEDA
;    APPLICANT:  Yoshikazu, SHIMADA
;    APPLICANT:  Kouichi, OZAKI
;    APPLICANT:  Sadahito, SIN
;    TITLE OF INVENTION:  hTFIIIA GENE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,376
;      FILING DATE:
;      CLASSIFICATION:  536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-523-376-3

  Query Match             17.7%;  Score 395.5;  DB 1;  Length 423;
  Best Local Similarity   30.8%;  Pred. No. 5.7e-29;
  Matches   98;  Conservative   56;  Mismatches   95;  Indels   69;  Gaps   16;

Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
              :|:|||      |::  ||         |:::||::  |   |::   |: ||  |: :|
Db         67 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 126

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
              |: |  : | ||| |  ||  ||::|: :||| |:   | :  |:: :||:| |  |
Db        127 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 186

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
               | : |:||:|:: | ||| :: |   : :|    | |  | | |  |||| :|   | |
Db        187 QKQYICSFEDCKKTFKKHQQMKIHQCQNTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 246

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
                | |   ||    :||: |  |::::| :
Db        247 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 301

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
              :||:  ||:      | ||:|| |::|   : : |::   || || |  |  |  :
Db        302 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 357

Qy        251 FHDGNIPDDLLKETEVKK 268
               ||   ||    : :|||
Db        358 VHD---PDKKKMKLKVKK 372
```

```
RESULT 6
US-09-949-016-6356
; Sequence 6356, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6356
;   LENGTH: 1191
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-6356

  Query Match             17.7%;  Score 393.5;  DB 4;  Length 1191;
  Best Local Similarity   35.8%;  Pred. No. 3.9e-28;
  Matches   98;  Conservative   40;  Mismatches  109;  Indels   27;  Gaps   12;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              :| | | |  | | ||:::|| |  | | |: ::|||  ::| ||| : | | || : |
Db        905 TREKPYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 960

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : :||: |  |||        |  |:| ||  | :||  | | :|| :  :| ||  :
Db        961 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1018

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
              ||   |::| | | | |||  ||: | |  | |:|:   | | | : | |  | |: | :
Db       1019 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1074

Qy        197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253
                 || :|||        |: |   |   | | : |  |  ||  |::  :  : | |
Db       1075 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1129

Qy        254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287
              |       |  : :|     :| | | |  :: |
Db       1130 GE------KPYKCEKCCKAFNQSSILTNHKKIHT 1157


RESULT 7
US-09-389-956-68
; Sequence 68, Application US/09389956
; Patent No. 6586579
; GENERAL INFORMATION:
;  APPLICANT: Huang, Shi
;  TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
;  TITLE OF INVENTION: Antibodies and Methods
;  FILE REFERENCE: P-LJ 3611
;  CURRENT APPLICATION NUMBER: US/09/389,956
;  CURRENT FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
;   LENGTH: 457
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-389-956-68

  Query Match             16.1%;  Score 358;  DB 4;  Length 457;
  Best Local Similarity   30.4%;  Pred. No. 2.3e-25;
  Matches   93;  Conservative   34;  Mismatches  105;  Indels   74;  Gaps   11;
```

```
Qy          14 ISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLE 73
               :||     |    |  : | | :   : |::|: |||  ||| | :  |:|:| |   :
Db         139 VSSLQEHRKIHEIFDCQECMKKFISANQLKRHMITHSEKRPYNCEI--CNKSFKRLDQVG 196

Qy          74 THIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT---------------------- 108
                | | ||| ||: | :||||   |   | |: ||:
Db         197 AHKVIHSEDKPYQCKLCGKGFAHRNVYKNHKKTHSEERPFQCDACKALFRTPFSLQRHLL 256

Qy         109 -----KSFKCTFENCQEAFYKHQSLRHHILSVHE--KTLTCKQCNKVFTRPSKLAQHKLK 161
                   ::|||   :|   | :  :| |   |||  |   |: ||| |  || |  ||
Db         257 IHNSERTFKC--HHCDATFKRKDTLNVHAQVVHERHKKYRCELCNKAFVTPSVLRSHKKS 314

Qy         162 H--------------------------HGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH 194
               |                          | |   |||  | | | |     |: ||: :|
Db         315 HTGEKEKVCPYCGQKFASSGTLRVHIRSHTGERPYQC--PYCEKGFSKNDGLKMHIR-TH 371

Qy         195 PKLK---CPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIF 251
                 : |   | :| |    |:||  | :|   |  | : || ||:  |: |  |: |  :
Db         372 TREKPYQCSECSKAFSQKRGLDEHKRTH---TGEKPFQCDVCDLA-FSLKKMLIRH-KMT 426

Qy         252 HDGNIP 257
               |: | |
Db         427 HNPNRP 432


RESULT 8
US-09-389-956-6
; Sequence 6, Application US/09389956
; Patent No. 6586579
; GENERAL INFORMATION:
;  APPLICANT: Huang, Shi
;  TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
;  TITLE OF INVENTION: Antibodies and Methods
;  FILE REFERENCE: P-LJ 3611
;  CURRENT APPLICATION NUMBER: US/09/389,956
;  CURRENT FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 93
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 647
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-389-956-6

  Query Match             16.1%;  Score 358;  DB 4;  Length 647;
  Best Local Similarity   30.4%;  Pred. No. 3.8e-25;
  Matches   93;  Conservative   34;  Mismatches  105;  Indels   74;  Gaps   11;

Qy          14 ISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLE 73
               :||     |    |   : | | :   : |::|: |||  ||| | :  |:|:| |   :
Db         324 VSSLQEHRKIHEIFDCQECMKKFISANQLKRHMITHSEKRPYNCEI--CNKSFKRLDQVG 381

Qy          74 THIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT---------------------- 108
                | | ||| ||: | :||||   |   | |: ||:
Db         382 AHKVIHSEDKPYQCKLCGKGFAHRNVYKNHKKTHSEERPFQCDACKALFRTPFSLQRHLL 441

Qy         109 -----KSFKCTFENCQEAFYKHQSLRHHILSVHE--KTLTCKQCNKVFTRPSKLAQHKLK 161
                   ::|||   :|   | :  :| |   |||  |   |: ||| |  || |  ||
Db         442 IHNSERTFKC--HHCDATFKRKDTLNVHAQVVHERHKKYRCELCNKAFVTPSVLRSHKKS 499

Qy         162 H--------------------------HGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH 194
               |                          | |   |||  | | | |     |: ||: :|
Db         500 HTGEKEKVCPYCGQKFASSGTLRVHIRSHTGERPYQC--PYCEKGFSKNDGLKMHIR-TH 556

Qy         195 PKLK---CPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIF 251
                 : |   | :| |    |:||  | :|   |  | : || ||:  |: |  |: |  :
Db         557 TREKPYQCSECSKAFSQKRGLDEHKRTH---TGEKPFQCDVCDLA-FSLKKMLIRH-KMT 611

Qy         252 HDGNIP 257
               |: | |
Db         612 HNPNRP 617


RESULT 9
US-09-949-016-10055
; Sequence 10055, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10055
;   LENGTH: 487
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10055

  Query Match             15.7%;  Score 351;  DB 4;  Length 487;
  Best Local Similarity   29.3%;  Pred. No. 1.2e-24;
  Matches  110;  Conservative   52;  Mismatches  155;  Indels   58;  Gaps   18;

Qy           6 ETKSISSLISSSSSSRPKKYICT----YEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDD 61
               : |:   ::|  |:|  :|    |     :: | |:: : | | ||   |: ::|| |  ::
Db         159 QCKASVKVVSKFSNSNKRKTRHTGEKHPKECGKSFQKFSHLTQHKVIHAGEKPYTC--EE 216

Qy          62 CDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQ 119
               | |||         |   |: :||| |  ||    :  |  :|:| ||  | :||  | |
Db         217 CGKAFKWSLIFNEHKRIHTGEKPFTCEECGSIFTTSSHFAKHKIIHTGEKPYKC--EECG 274

Qy         120 EAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFK 178
               :|| :  :| :|  :    || :||::|  |:||  |  |:||  | |  | |:|:    | |
Db         275 KAFNRFTTLTKHKRIHAGEKPITCEECRKIFTSSSNFAKHKRIHTGEKP-YKCEE--CGK 331

Qy         179 NFQTWSVLQFHIKQSHPKLK---CPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV 235
                 |   : |  | |: |    |  | :|||        |: |   |   |  : : || |
Db         332 AFNRSTTLTKH-KRIHTGEKPYTCEECGKAFRQSSKLNEHKKVH---TGERPYKCDEC-- 385

Qy         236 GK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEE 294
               || | :   | ||  | | |  |     |  |  |         |:   :|  : :  | :|
Db         386 GKAFGRSRVLNEHKKI-HTGEKP---YKCEECGKAFRRSTDRSQHKKIHSAD-KPYKCKE 440

Qy         295 DEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDR 354
                                          :: | |:       |  | |    |   |  :||:
Db         441 ----------------------CDKAFKQFSL-----LSQHKKIHTVDKPYKC--KDCDK 471

Qy         355 MFSREYDLRRHLKWH 369
                | |    | :| | |
Db         472 AFKRFSHLNKHKKIH 486


RESULT 10
US-09-949-016-8627
; Sequence 8627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
```

```
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8627
;   LENGTH: 685
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-8627

  Query Match             15.7%;  Score 350;  DB 4;  Length 685;
  Best Local Similarity   28.8%;  Pred. No. 2.4e-24;
  Matches  110;  Conservative   60;  Mismatches  160;  Indels   52;  Gaps   19;

Qy          6 ETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKA 65
              :  |:      |  | :|  | |    | ||: | | | || | || ::|| |:  :| ||
Db        307 QNSSLKKHQKSHMSEKP--YEC--NECGKAFRRSSNLIQHQRIHSGEKPYVCS--ECGKA 360

Qy         66 FFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFY 123
              | | |:|  |  :|: :||| |  |||  :   ||::|:  ||  | ::|   :| : |
Db        361 FRRSSNLIKHHRTHTGEKPFECGECGKAFSQSAHLRKHQRVHTGEKPYEC--NDCGKPFS 418

Qy        124 KHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQT 182
              :  :| :|| :   ||   | | | |:: | | ||:  | |  | : |:   | | |
Db        419 RVSNLIKHHRVHTGEKPYKCSDCGKAFSQSSSLIQHRRIHTGEKP-HVCN--VCGKAFSY 475

Qy        183 WSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FA 239
              |||: |  |       :|  |||       |  |   |   |  | : |  |  || |
Db        476 SSVLRKHQIIHTGEKPYRCSVCGKAFSHSSALIQHQGVH---TGDKPYACHEC--GKTFG 530

Qy        240 KKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDE 299
              : : |: |  : | |  | :  :  :     : | |  :::|         ||  |  :
Db        531 RSSNLILHQRV-HTGEKPYECTECGKTFSQSSTLIQHQRIHN-------GLKPHECNQC- 581

Qy        300 EDSLDEKRSDVRSDSMSAQRSI----KSFTASLEGSKSVS--------KLISNSGKKINC 347
                  ::  || ::   : :    | :|  :|  |  |        ::|    :   |
Db        582 ------GKAFNRSSNLIHHQKVHTGEKPYTC-VECGKGFSQSSHLIQHQIIHTGERPYKC 634

Qy        348 PKNNCDRMFSREYDLRRHLKWH 369
                : | : ||:   | :| : |
Db        635 --SECGKAFSQRSVLIQHQRIH 654


RESULT 11
US-09-949-016-9434
; Sequence 9434, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9434
;   LENGTH: 639
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-9434

  Query Match             15.4%;  Score 344;  DB 4;  Length 639;
  Best Local Similarity   27.5%;  Pred. No. 7.9e-24;
  Matches   98;  Conservative   47;  Mismatches  129;  Indels   82;  Gaps   15;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |    | ||:     |  ||| |: ::||:|   :| | |   |:|  |: ||: :
Db        356 KPYECIQ--CGKAHGHKHALTDHLRIHTGEKPYECA--ECGKTFRHSSNLIQHVRSHTGE 411

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQSLRHHILSVH--EKTL 140
              ||: |  |||        |  |  |||          | :||    ||  | : :|  ||
Db        412 KPYECKECGKSFRYNSSLTEHVRTHTGEIPYECNECGKAFKYSSSLTKH-MRIHTGEKPF 470

Qy        141 TCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIK--QSHPKLK 198
              | :| | |:: | |  |:  |     | |:|:   | | |   | | :|:: :
Db        471 ECNECGKAFSKKSHLIIHQRTHTKEKP-YKCNE--CGKAFGHSSSLTYHMRTHTGESPFE 527

Qy        199 CPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIP 257
              | :||||    :||: |   |   | | : |: |  || |::|: |: |   | |
Db        528 CNQCGKGFKQIEGLTQHQRVH---TGEKPYECNEC--GKAFSQKSHLIVHQRT-HTG--- 578

Qy        258 DDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSA 317
                                                  |:  | :  ||  : :|  :
Db        579 -----------------------------------EKPYECNECEKAFNAKSQLVIH 600

Qy        318 QRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWHDDN 372
              ||                |::|:|   |  | | : | :   | :|:| | ::
Db        601 QR------------------SHTGEKPYEC--NECGKTFKQNASLTKHVKTHSED 635


RESULT 12
US-09-949-016-11494
; Sequence 11494, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11494
;   LENGTH: 746
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-11494

  Query Match             15.4%;  Score 343;  DB 4;  Length 746;
  Best Local Similarity   27.3%;  Pred. No. 1.2e-23;
  Matches  118;  Conservative   65;  Mismatches  141;  Indels  108;  Gaps   22;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | : ||    | ||:   | | || | |: ::||:|  ::| ||| : | |  |   |: :
Db        336 KPFKCT--ECGKAFRLSSKLIQHQRIHTGEKPYRC--EECGKAFGQSSSLIHHQRIHTGE 391

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH--------- 131
              :|: |  |||  : :  | ||: |||  : : |  : | :|| :  :| |
Db        392 RPYGCRECGKAFSQQSQLVRHQRTHTGERPYPC--KECGKAFSQSSTLAQHQRMHTGEKA 449

Qy        132 -ILS------------VH--EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176
               ||            :|  ||   | :| | |   |:|:||:| | |  | |:|:   |
Db        450 QILKASDSPSLVAHQRIHAVEKPFKCDECGKAFRWISRLSQHQLIHTGEKP-YKCNK--C 506

Qy        177 FKNFQTWSVLQFHIKQSH---PKLKCPKCGKGCVGKKGLSSHMLSHDD-STMIKIWTCDY 232
              | |   | |  | :::|      || :|||| |      ||:: |    | | : |:
Db        507 TKAFGCSSRLIRH-QRTHTGEKPFKCDECGKGFV----QGSHLIQHQRIHTGEKPYVCND 561

Qy        233 CDVGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLL---------DQGSKLNNL 282
              |  || |:: : |: |  | | |  | : |:  :   :   |          ::  | |
Db        562 C--GKAFSQSSSLIYHQRI-HKGEKPYECLQCGKAFSMSTQLTIHQRVHTGERPYKCNEC 618
```

```
Qy        283 ------------HELETEKLKVEEDEE------DEEDSLDEKRSDVRSDSMSAQRSIKSF 324
                          |::    :|  | |          :: :|   |:      :   :
Db        619 GKAFSQNSTLFQHQIIHAGVKPYECSECGKAFSRSSYLIEHQRIHTRA------QWFYEY 672

Qy        325 TASLEGSKSVSKLISNSGKKI----NCPK----------------------NNCDRMFS 357
                :|||| ||:   |: ||:    :| |                      |:| : |:
Db        673 GNALEGSTFVSRKKVNTIKKLHQCEDCEKIFRWRSHLIIHQRIHTGEKPYKCNDCGKAFN 732

Qy        358 REYDLRRHLKWH 369
              |   | :| | |
Db        733 RSSRLTQHQKIH 744


RESULT 13
US-09-949-016-8149
; Sequence 8149, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8149
;   LENGTH: 769
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-8149

  Query Match             15.4%;  Score 343;  DB 4;  Length 769;
  Best Local Similarity   32.2%;  Pred. No. 1.3e-23;
  Matches   86;  Conservative   35;  Mismatches  100;  Indels   46;  Gaps   11;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  | | | ::  |  : | | |: ::|::|  |  | | | : |: : |   |: :
Db        516 KPYKC--EECGKGFSSASSFQSHQRVHTGEKPFRCNV--CGKGFSQSSYFQAHQRVHTGE 571

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSVH--EK 138
              ||: | ||||  |   :|  |:  ||  | :||  | | : | :  :|: |  |||  ||
Db        572 KPYKCEVCGKRFNWSLNLHNHQRVHTGEKPYKC--EECGKGFSQASNLQAH-QSVHTGEK 628

Qy        139 TLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPK 196
                | | | |:: | |  |:  | |  | |:||   | | |   | || |  |
Db        629 PFKCDACQKRFSQASHLQAHQRVHTGEKP-YKCD--TCGKAFSQRSNLQVHQIIHTGEKP 685

Qy        197 LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVG-------------------- 236
               || :|||      |||:|   |   |  | :||  |  |
Db        686 FKCEECGKEFSWSAGLSAHQRVH---TGEKPYTCQQCGKGFSQASHFHTHQRVHTGERPY 742

Qy        237 -------KFAKKNELVEHYNIFHDGNI 256
                    |:::: |: |  :   ||:
Db        743 ICDVCCKGFSQRSHLIYHQRVHTGGNL 769


RESULT 14
US-09-538-092-1275
; Sequence 1275, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1275
;   LENGTH: 810
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number Q06730
US-09-538-092-1275

  Query Match             15.4%;  Score 343;  DB 4;  Length 810;
  Best Local Similarity   28.3%;  Pred. No. 1.4e-23;
  Matches  112;  Conservative   57;  Mismatches  153;  Indels   74;  Gaps   20;

Qy         18 SSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIV 77
              | :  | : |    | ||::  | |  | |||: ::||:|    | | | :|| |  |
Db        377 SHTGEKPFEC--NECGKAFSHKSALTLHQRTHTGEKPYQCNA--CGKTFCQKSDLTKHQR 432

Qy         78 SHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILS 134
              :|:  ||: |  |||       ||| |: |||  | |:|    | ::| :  :| :|  :
Db        433 THTGLKPYECYECGKSFRVTSHLKVHQRTHTGEKPFECL--ECGKSFSEKSNLTQHQRIH 490

Qy        135 VHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH 194
              : :|:  |  | | |   | | :|:: | |  | |:|    | | |   | | | :::|
Db        491 IGDKSYECNACGKTFYHKSLLTRHQIIHTGWKP-YECYE--CGKTFCLKSDLTVH-QRTH 546

Qy        195 PKLK---CPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
                 |   ||:|||    |  || |  :|   |  | : |  |  || |  |: | :| |
Db        547 TGQKPFACPECGKFFSHKSTLSQHYRTH---TGEKPYECHEC--GKIFYNKSYLTKH-NR 600

Qy        251 FHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEE-----------D 298
              | |       |  |  :      | |:|       |:  ::   |:
Db        601 THTGE------KPYECNECGKAFYQKSQL-------TQHQRIHIGEKPYKCNECGKAFCH 647

Qy        299 EEDSLDEKRSDVRSDSMSAQRSIKSFTASL---------EGSKSV-----------SKL 337
              :   :  :|:  :           |||             | |              | |
Db        648 KSALIVHQRTHTQEKPYKCNECGKSFCVKSGLIFHERKHTGEKPYECNECGKFFRHKSSL 707

Qy        338 I----SNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                   :::|:| :|  | | ::| |: :| :| : |
Db        708 TVHHRAHTGEK-SCQCNECGKIFYRKSELAQHQRSH 742


RESULT 15
US-09-538-092-1195
; Sequence 1195, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1195
;   LENGTH: 469
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P52742
US-09-538-092-1195

  Query Match             15.1%;  Score 337;  DB 4;  Length 469;
  Best Local Similarity   28.2%;  Pred. No. 2.3e-23;
  Matches  103;  Conservative   49;  Mismatches  163;  Indels   50;  Gaps   17;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | ||   | : :|: : | || |||: ::||:|:  :| |:|  :|    |  :|: :
Db         79 KPYKCTQ--CGRTFNQIAPLIQHQRTHTGEKPYECS--ECGKSFSFRSSFSQHERTHTGE 134

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHEKT 139
              ||: || |||      || :|   ||  | ::|    | :||    || :|  :   ||
Db        135 KPYECSECGKAFRQSIHLTQHLRIHTGEKPYQC--GECGKAFSHSSSLTKHQRIHTGEKP 192

Qy        140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
                | :| | ||: : | ||:  | |  | |:|    | | |   ::|  |  |
Db        193 YECHECGKAFTQITPLIQHQRTHTGEKP-YECGE--CGKAFSQSTLLTEHRRIHTGEKPY 249

Qy        198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNI 256
               | :|||       || |  :|   |  | : |  |  || | :   | :|  | | |
Db        250 GCNECGKTFSHSSSLSQHERTH---TGEKPYECSQC--GKAFRQSTHLTQHQRI-HTGE- 302

Qy        257 PDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----------EDEEDSLD 304
                |  |           | |     : | :   | ::           : :
Db        303 -----KPYECNDCGKAFSHSSSLTKHQRIHTGEKPYECNQCGRAFSQLAPLIQHQRIHTG 357

Qy        305 EKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRR 364
              ||  :      :: |:   |  :|  :      |    |   | | : ||     | :
Db        358 EKPYECNQCGRASARA----TLLIEHQR-----IHTKEKPYGC--NECGKSFSHSSSLSQ 406

Qy        365 HLKWH 369
              | : |
Db        407 HERTH 411


Search completed: February  8, 2005, 15:23:40
Job time : 45 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:22:12 ; Search time 129 Seconds
(without alignments)
1040.228 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1373511 seqs, 325702437 residues

Total number of hits satisfying chosen parameters: 1373511

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2225	99.8	412	14	US-10-032-585-7493	Sequence 7493, Ap
2	445	20.0	564	14	US-10-128-714-3313	Sequence 3313, Ap
3	445	20.0	564	14	US-10-128-714-8313	Sequence 8313, Ap
4	400	17.9	404	15	US-10-108-260A-4601	Sequence 4601, Ap
5	398	17.9	487	14	US-10-029-386-33849	Sequence 33849, A
6	397.5	17.8	423	16	US-10-734-564-83	Sequence 83, Appl
7	395.5	17.7	409	14	US-10-106-698-5955	Sequence 5955, Ap
8	395.5	17.7	1191	16	US-10-408-765A-2105	Sequence 2105, Ap
9	392.5	17.6	595	16	US-10-203-052B-8	Sequence 8, Appli
10	392	17.6	555	15	US-10-094-749-2331	Sequence 2331, Ap
11	391	17.5	592	15	US-10-108-260A-3669	Sequence 3669, Ap
12	390	17.5	540	15	US-10-104-047-3748	Sequence 3748, Ap
13	390	17.5	576	15	US-10-094-749-2365	Sequence 2365, Ap
14	388	17.4	568	16	US-10-203-052B-7	Sequence 7, Appli
15	387.5	17.4	742	14	US-10-029-386-31985	Sequence 31985, A
16	387.5	17.4	839	16	US-10-408-765A-1850	Sequence 1850, Ap
17	385	17.3	577	15	US-10-381-327-3	Sequence 3, Appli
18	378	17.0	613	16	US-10-408-765A-1517	Sequence 1517, Ap
19	376	16.9	499	15	US-10-108-260A-2954	Sequence 2954, Ap
20	374	16.8	566	15	US-10-322-774-10	Sequence 10, Appl
21	373	16.7	452	16	US-10-471-450-5	Sequence 5, Appli
22	372.5	16.7	586	15	US-10-104-047-2592	Sequence 2592, Ap
23	372	16.7	393	10	US-09-911-261A-12	Sequence 12, Appl
24	372	16.7	393	14	US-10-057-408-12	Sequence 12, Appl
25	372	16.7	393	15	US-10-333-487-12	Sequence 12, Appl
26	371	16.6	403	14	US-10-029-386-32933	Sequence 32933, A
27	370	16.6	530	15	US-10-094-749-2723	Sequence 2723, Ap
28	369.5	16.6	464	15	US-10-108-260A-3590	Sequence 3590, Ap
29	367	16.5	514	14	US-10-029-386-34112	Sequence 34112, A
30	367	16.5	610	9	US-09-989-920-212	Sequence 212, App
31	367	16.5	641	15	US-10-363-616-421	Sequence 421, App
32	364.5	16.4	444	15	US-10-108-260A-3223	Sequence 3223, Ap
33	364	16.3	516	9	US-09-974-298-118	Sequence 118, App
34	364	16.3	535	15	US-10-264-049-3016	Sequence 3016, Ap
35	363	16.3	516	15	US-10-094-749-1976	Sequence 1976, Ap
36	361	16.2	803	16	US-10-408-765A-1049	Sequence 1049, Ap
37	360	16.2	529	15	US-10-108-260A-3085	Sequence 3085, Ap
38	360	16.2	791	16	US-10-471-450-18	Sequence 18, Appl
39	360	16.2	924	14	US-10-007-280A-231	Sequence 231, App
40	359.5	16.1	365	15	US-10-094-749-2297	Sequence 2297, Ap
41	359.5	16.1	427	16	US-10-755-889-370	Sequence 370, App
42	358.5	16.1	473	14	US-10-029-386-32047	Sequence 32047, A
43	358	16.1	406	9	US-09-764-864-976	Sequence 976, App
44	358	16.1	457	15	US-10-377-079-68	Sequence 68, Appl
45	358	16.1	647	15	US-10-377-079-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-032-585-7493
; Sequence 7493, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7493
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7493

Query Match 99.8%; Score 2225; DB 14; Length 412;
Best Local Similarity 99.8%; Pred. No. 2.3e-154;
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVE 60

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
```

```
Qy        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240

Qy        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360

Qy        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
```

RESULT 2
US-10-128-714-3313
; Sequence 3313, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3313
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3313

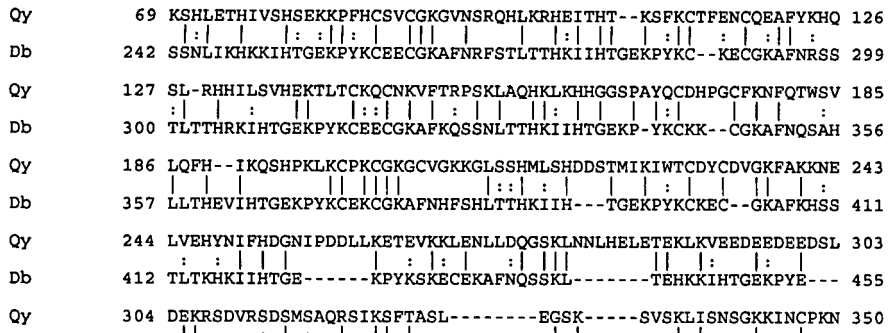
```
  Query Match             20.0%;  Score 445;  DB 14;  Length 564;
  Best Local Similarity   30.3%;  Pred. No. 3.9e-24;
  Matches  125;  Conservative   62;  Mismatches  164;  Indels   62;  Gaps   21;

Qy         14 ISSSSSSRP---KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
              ||::|:   |    | : | ::|| ||:|||: |::|||:|:|:| :||| ::|||  | | |
Db         71 ISTTSAKYPSELKTHRCPFDGCTKAFNRPARLQEHLRSHNNERIFKCTFEECDKTFLRAS 130

Qy         71 HLETHIVS-HSEKKPFHCS--VCGKGVNSRQHLKRHEITH--TKSFKCT-FENCQEAFYK 124
              || || | |: : :  |     |||   :   |:||   |       ::|| :  | | | |
Db        131 HLNHHIKSAHTGVRDYVCDRPGCGKSFVTGSRLRRHLAAHDGRDKYRCTEYPPCNETFRK 190

Qy        125 HQSLRHHILSVH--EKTLTC--------KQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH- 173
              | :|: ||:: |  :|   |        ::|  |      |  |: : |        :  |
Db        191 HSTLQKHIMTAHLKQKPFQCPHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFSCTEC 249

Qy        174 ----PGCFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHM-LSHDDSTM--IK 226
                  |   |:::|| ||: ||  :|| |   |   |:  ::| | :   |
Db        250 SQHAEGAEATFPTYALLQAHIRSVHPP-QCPNCALTCATSRELRRHLEVAHGDVSLEERK 308

Qy        227 IWTCDY--CDVGKFAKKNELVEHYNIFHDGNI-----PDDLLKETEVKKLENLLDQ---- 275
              |: |   ||  | ||  |   |   | |           ||      :|     |  |
Db        309 IFPCTVPGCD-RSFTKKGNLTVHIRTVHQGEKRFVCGETDLSSSKKVSGWNN--DNGCGK 365

Qy        276 --GSKLNNLHELETEKLKVEEDEEDEEDSL----DEKRSDVRSDSMSAQRSIKSFTASLE 329
                ||||     : |  |  :  : :    |     | : |  |  :||        |:|
Db        366 RYGSKLALEEHIRTAHLGYQNAKAERRQRLGITRDRQHSTATSPGVSA-------LAALT 418

Qy        330 GSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHL--KWHDDNLQRIESFL 380
              |    :     :|: |   :|    | |:|| |:  |  |    :  : ||
Db        419 GEGYAEE----TGRHIACLVESCPHRFHRDYDLWVHMSGKHHFSEEETRDLFL 467
```

RESULT 3
US-10-128-714-8313
; Sequence 8313, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8313
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8313

```
  Query Match             20.0%;  Score 445;  DB 14;  Length 564;
  Best Local Similarity   30.3%;  Pred. No. 3.9e-24;
  Matches  125;  Conservative   62;  Mismatches  164;  Indels   62;  Gaps   21;

Qy         14 ISSSSSSRP---KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
              ||::|:   |    | : | ::|| ||:|||: |::|||:|:|:| :||| ::|||  | | |
Db         71 ISTTSAKYPSELKTHRCPFDGCTKAFNRPARLQEHLRSHNNERIFKCTFEECDKTFLRAS 130

Qy         71 HLETHIVS-HSEKKPFHCS--VCGKGVNSRQHLKRHEITH--TKSFKCT-FENCQEAFYK 124
              || || | |: : :  |     |||   :   |:||   |       ::|| :  | | | |
Db        131 HLNHHIKSAHTGVRDYVCDRPGCGKSFVTGSRLRRHLAAHDGRDKYRCTEYPPCNETFRK 190

Qy        125 HQSLRHHILSVH--EKTLTC--------KQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH- 173
              | :|: ||:: |  :|   |        ::|  |      |  |: : |        :  |
Db        191 HSTLQKHIMTAHLKQKPFQCPHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFSCTEC 249

Qy        174 ----PGCFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHM-LSHDDSTM--IK 226
                  |   |:::|| ||: ||  :|| |   |   |:  ::| | :   |
Db        250 SQHAEGAEATFPTYALLQAHIRSVHPP-QCPNCALTCATSRELRRHLEVAHGDVSLEERK 308

Qy        227 IWTCDY--CDVGKFAKKNELVEHYNIFHDGNI-----PDDLLKETEVKKLENLLDQ---- 275
              |: |   ||  | ||  |   |   | |           ||      :|     |  |
Db        309 IFPCTVPGCD-RSFTKKGNLTVHIRTVHQGEKRFVCGETDLSSSKKVSGWNN--DNGCGK 365

Qy        276 --GSKLNNLHELETEKLKVEEDEEDEEDSL----DEKRSDVRSDSMSAQRSIKSFTASLE 329
                ||||     : |  |  :  : :    |     | : |  |  :||        |:|
Db        366 RYGSKLALEEHIRTAHLGYQNAKAERRQRLGITRDRQHSTATSPGVSA-------LAALT 418
```

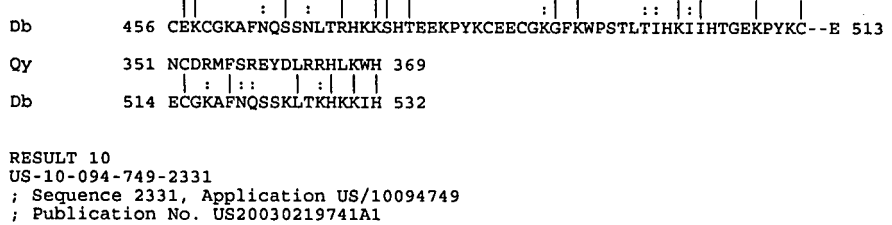
```
Qy        330 GSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHL--KWHDDNLQRIESFL 380
Db        419 GEGYAEE----TGRHIACLVESCPHRFHRDYDLWVHMSGKHHFSEEETRDLFL 467


RESULT 4
US-10-108-260A-4601
; Sequence 4601, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4601
;   LENGTH: 404
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-4601

  Query Match            17.9%;  Score 400;  DB 15;  Length 404;
  Best Local Similarity  33.9%;  Pred. No. 5e-21;
  Matches  121;  Conservative   39;  Mismatches  163;  Indels   34;  Gaps   19;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db         39 KPYKC--EECGKAFNRSSILTKHKIIHTGEKPYKC--EECGKGFSSVSTLNTHKAIHAEE 94

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSVH--EK 138
Db         95 KPYKCEECGKASNSSSKLMEHKRIHTGEKPYKC--EECGKAFSWSSSLTEH-KRIHAGEK 151

Qy        139 TLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPK 196
Db        152 PYKCEECGKAFNRSSILTKHKIIHTGEKP-YKCE--GCGKAFSKVSTLNTHKAIHAEEKP 208

Qy        197 LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGN 255
Db        209 YKCEECGKASNSSSKLMEHKRIH---TGEKPYKCEEC--GKAFSWSSSLTEHKRI-HAGE 262

Qy        256 IPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEE-DEEDEEDSLDEKRSDVRSDS 314
Db        263 KP---YKCEECGKAFTWSSSFTKHKRIHAAE-KPYKCEECGKGFSTFSILTKHKIIHTGE 318

Qy        315 --MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
Db        319 KRYKCEECGKAF--SWSSILTEHKIIHTGEKPYKC--EECGKAFSRSSSLTRHKRIH 371


RESULT 5
US-10-029-386-33849
; Sequence 33849, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33849
;   LENGTH: 487
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC011467.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.67
;   OTHER INFORMATION: SWISSPROT HIT: Q03923, EVALUE 0.00e+00
US-10-029-386-33849

  Query Match            17.9%;  Score 398;  DB 14;  Length 487;
  Best Local Similarity  30.2%;  Pred. No. 8.7e-21;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
Db        120 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 175

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
Db        176 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 233

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
Db        234 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 290

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
Db        291 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 345

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
Db        346 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 391

Qy        315 MSAQRSI-------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
Db        392 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 447

Qy        362 LRRHLKWHD--------------DNLQRIESFLNSIEKEE 387
Db        448 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 487


RESULT 6
US-10-734-564-83
; Sequence 83, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
;  APPLICANT: Christopher C Burgess et al
;  TITLE OF INVENTION: Detection Methods Using TIMP1
;  FILE REFERENCE: 1657/2012
;  CURRENT APPLICATION NUMBER: US/10/734,564
;  CURRENT FILING DATE:  2003-12-12
;  NUMBER OF SEQ ID NOS: 138
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 83
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-734-564-83

  Query Match            17.8%;  Score 397.5;  DB 16;  Length 423;
  Best Local Similarity  31.1%;  Pred. No. 8e-21;
  Matches   99;  Conservative   55;  Mismatches   95;  Indels   69;  Gaps   16;

Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
Db         67 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 126

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
Db        127 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 186

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
Db        187 QKQYICSFEDCKKTFKKHQQLKIHQCQNTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 246

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
```

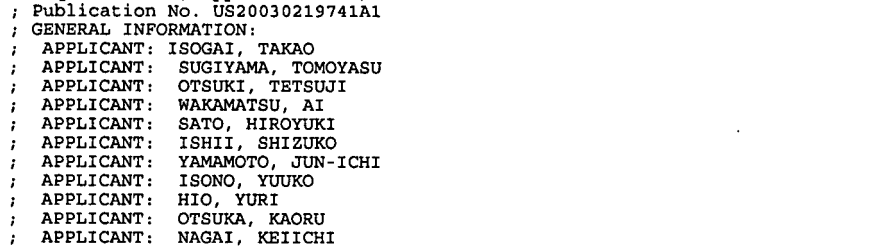
```
Db        247 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 301

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250

Db        302 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 357

Qy        251 FHDGNIPDDLLKETEVKK 268

Db        358 VHD---PDKKKMKLKVKK 372


RESULT 7
US-10-106-698-5955
; Sequence 5955, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5955
;   LENGTH: 409
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (3)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5955

  Query Match             17.7%;  Score 395.5;  DB 14;  Length 409;
  Best Local Similarity   31.1%;  Pred. No. 1.1e-20;
  Matches   99;  Conservative   54;  Mismatches   96;  Indels   69;  Gaps   16;

Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53

Db         53 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 112

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107

Db        113 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 172

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164

Db        173 QKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 232

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196

Db        233 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 287

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250

Db        288 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 343

Qy        251 FHDGNIPDDLLKETEVKK 268

Db        344 VHD---PDKKKMKLKVKK 358


RESULT 8
US-10-408-765A-2105
; Sequence 2105, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
;  APPLICANT: Ghosh, Soumitra S.
;  APPLICANT:  Fahy, Eoin D.
;  APPLICANT:  Zhang, Bing
;  APPLICANT:  Gibson, Bradford W.
;  APPLICANT:  Taylor, Steven W.
;  APPLICANT:  Glenn, Gary M.
;  APPLICANT:  Warnock, Dale E.
;  TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
;  TITLE OF INVENTION:  IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
;  FILE REFERENCE: 660088.465
;  CURRENT APPLICATION NUMBER: US/10/408,765A
;  CURRENT FILING DATE:  2003-04-04
;  NUMBER OF SEQ ID NOS: 3077
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2105
;   LENGTH: 1191
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-408-765A-2105

  Query Match             17.7%;  Score 395.5;  DB 16;  Length 1191;
  Best Local Similarity   35.8%;  Pred. No. 4e-20;
  Matches   98;  Conservative   40;  Mismatches  109;  Indels   27;  Gaps   12;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79

Db        905 TREKTYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 960

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136

Db        961 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1018

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196

Db       1019 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1074

Qy        197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253

Db       1075 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1129

Qy        254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287

Db       1130 GE------KPYKCEKCGKAFNQSSILTNHKKIHT 1157


RESULT 9
US-10-203-052B-8
; Sequence 8, Application US/10203052B
; Publication No. US20040109854A1
; GENERAL INFORMATION:
;  APPLICANT: KOMED CO., LTD.
;  APPLICANT:  Lee, Zang Hee
;  APPLICANT:  Kim, Hong Hee
;  TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein
;  FILE REFERENCE: 59520-00002
;  CURRENT APPLICATION NUMBER: US/10/203,052B
;  CURRENT FILING DATE:  2003-11-03
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-203-052B-8

  Query Match             17.6%;  Score 392.5;  DB 16;  Length 595;
  Best Local Similarity   31.7%;  Pred. No. 2.8e-20;
  Matches  120;  Conservative   48;  Mismatches  160;  Indels   51;  Gaps   17;

Qy         10 ISSLISSSS-SSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFR 68

Db        186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTKHKRIHTCEKPYKC--EECGKAFNQ 241
```



RESULT 10
US-10-094-749-2331
; Sequence 2331, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2331
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2331

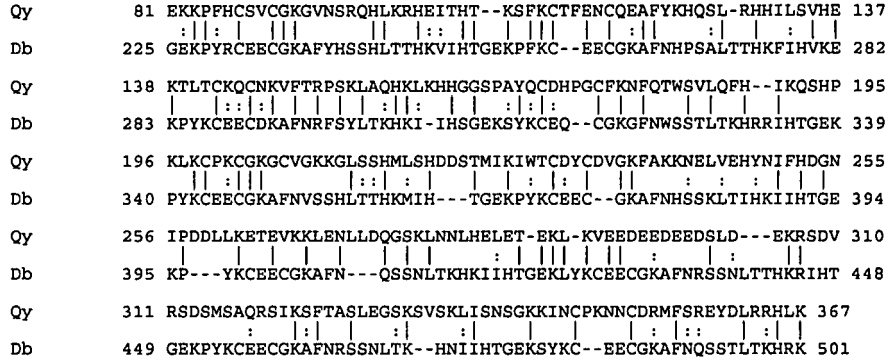
Query Match 17.6%; Score 392; DB 15; Length 555;
Best Local Similarity 32.5%; Pred. No. 2.8e-20;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17;

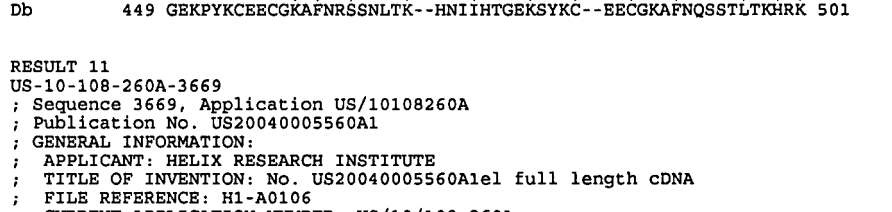


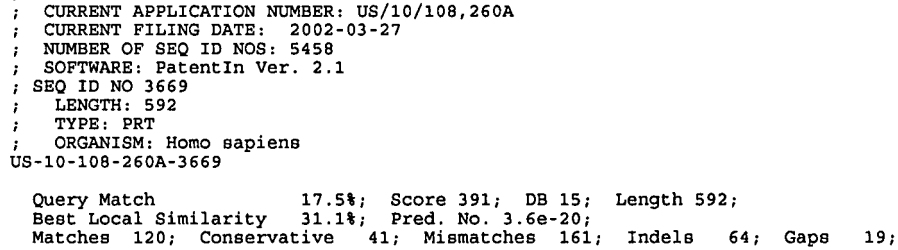


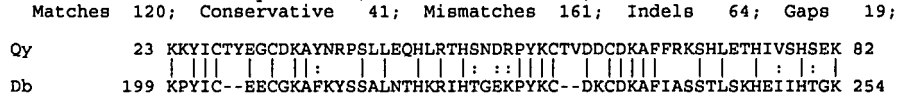
RESULT 11
US-10-108-260A-3669
; Sequence 3669, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3669
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3669

Query Match 17.5%; Score 391; DB 15; Length 592;
Best Local Similarity 31.1%; Pred. No. 3.6e-20;
Matches 120; Conservative 41; Mismatches 161; Indels 64; Gaps 19;









RESULT 12

```
US-10-104-047-3748
; Sequence 3748, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3748
;   LENGTH: 540
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3748

  Query Match            17.5%;  Score 390;  DB 15;  Length 540;
  Best Local Similarity   30.0%;  Pred. No. 3.8e-20;
  Matches  107;  Conservative   61;  Mismatches  155;  Indels   34;  Gaps   15;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        198 KLYKC--QECDRTFNQFSNLTEYKKDYAREKPYKC--EECGKAFNQSSHLTTHKIIHTGE 253

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQSL-RHHILSVHEKTLT 141
Db        254 KPYKCEECGKAFNQFSNLTTHKKIHTGEQPYICEECGKAFTQSSTLTTHKRIHTGEKPYK 313

Qy        142 CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKLKC 199
Db        314 CEECGKAFNRSSKLTEHKNIHTGEQP-YKCEE--CGKAFNRSSNLTEHRKIHTEEKPYKC 370

Qy        200 PKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPD 258
Db        371 KECGKAFKHSSALTTHKRIH---TGEKPYKCEEC--GKAFNRSSKLTEHKKL-HTGK--- 421

Qy        259 DLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----EDEEDSLDEKRSDVRSD 313
Db        422 ---KPYKCEECGKAFIQSSKLTEHKKIHSGEIPYKCEECGKAFKHSSSLTTHKRIHTGEK 478

Qy        314 SMSAQRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWH 369
Db        479 PYKCEECGKAFSRS---SKLTEHKIIHTGEKPYKCER--CDKAFNQSANLTKHKKIH 530


RESULT 13
US-10-094-749-2365
; Sequence 2365, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2365
;   LENGTH: 576
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-749-2365

  Query Match            17.5%;  Score 390;  DB 15;  Length 576;
  Best Local Similarity   28.6%;  Pred. No. 4.1e-20;
  Matches  119;  Conservative   57;  Mismatches  168;  Indels   72;  Gaps   17;

Qy          3 ESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDC 62
Db        217 ECDKAFNQSSTLTTHKIIHTREKLNEYKECGKAFNQSSHLTRHKIIHTGEKPYKC--EEC 274

Qy         63 DKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQE 120
Db        275 GKAFNQSSHLTRHKIIHTGEKPYKCEECGKAFRQSSHLTTHKIIHTGEKPYKC--EECGK 332

Qy        121 AFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKN 179
Db        333 AFNKSSHLTRHKSIHTGEKPYQCEKCGKASNQSSNLTEHKNIHTEEKP-YKCEE--CGKA 389

Qy        180 FQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK 237
Db        390 FNQFSNLTTHKRIHTGEKPYKCEECGKAFNQSSILTTHKRIH---TGEKSYKCEEC--GK 444

Qy        238 -FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE 296
Db        445 AFYRSSKLTEHKKI-HTGEKP-----------YTCEECGKAFNHSSHLATHKV-IHTGE 490

Qy        297 EDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMF 356
Db        491 KPYQ---------------CEECGKAFNQSSHLTR--HKRIHTGEKPYQCEK--CGKAF 530

Qy        357 SREYDLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
Db        531 NQSSNLTGHKKIH-------------------TGEKLYKPKRCNSDFENTSKFSK 566


RESULT 14
US-10-203-052B-7
; Sequence 7, Application US/10203052B
; Publication No. US20040109854A1
; GENERAL INFORMATION:
;  APPLICANT: KOMED CO., LTD.
;  APPLICANT:  Lee, Zang Hee
;  APPLICANT:  Kim, Hong Hee
;  TITLE OF INVENTION: A Novel TRAF6 Inhibiting Protein
;  FILE REFERENCE: 59520-00002
;  CURRENT APPLICATION NUMBER: US/10/203,052B
;  CURRENT FILING DATE:  2003-11-03
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 568
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-203-052B-7

  Query Match            17.4%;  Score 388;  DB 16;  Length 568;
  Best Local Similarity   28.5%;  Pred. No. 5.7e-20;
  Matches  105;  Conservative   62;  Mismatches  167;  Indels   34;  Gaps   13;

Qy         15 SSSSSSRPKKYIC----TYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
```

```
Db        212 SSKLTKHKRIYTCEKLYKCQECDRTFNQFSNLTEYKKDYAREKPYKC--EECGKAFNQSS 269

Qy         71 HLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQSL-R 129

Db        270 HLTTHKIIHTGEKPYKCEECGKASNQFSNLTTHKKIHTGEQPYICEECGKAFTQSSTLTT 329

Qy        130 HHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH 189

Db        330 HKRIHTGEKPYKCEECGKAFNRSSKLTEHKNIHTGEQP-YKCEE--CGKAFNRSSNLTEH 386

Qy        190 --IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEH 247

Db        387 RKIHTEEKPYKCKECGKAFKHSSALTTHKRIH---TGEKPYKCEEC--GKAFNRSSKLTE 441

Qy        248 YNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----EDEEDS 302

Db        442 YKKLHTGK------KPYKCEECGKAFIQSSKLTEHKKIHSGEIPYKCEECGKAFKHSSSL 495

Qy        303 LDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYD 361

Db        496 TTHKRIHTGEKPYKCEECSKAFSRS---SKLTEHKIIHTGEKPYECER--CDKAFNQSAN 550

Qy        362 LRRHLKWH 369

Db        551 LTKHKKIH 558
```

RESULT 15
US-10-029-386-31985
; Sequence 31985, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31985
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010328.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: SWISSPROT HIT: Q05481, EVALUE 0.00e+00
US-10-029-386-31985

Query Match 17.4%; Score 387.5; DB 14; Length 742;
Best Local Similarity 30.7%; Pred. No. 8.6e-20;
Matches 112; Conservative 46; Mismatches 154; Indels 53; Gaps 16;

```
Qy         11 SSL-ISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRK 69

Db        289 SSLAIHQATHSGEKPYKC--NECGKVFTQNSHLTNHWRIHTGEKPYKC--NECGKAFGVR 344

Qy         70 SHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQS 127

Db        345 SSLAIHLVIHTGEKPYKCHECGKVFRRNSHLARHQLIHTGEKPYKC--NECGKAFRAHSN 402

Qy        128 L-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVL 186

Db        403 LTTHQVIHTGEKPYKCNECGKVFTQNSHLANHQRIHTGVKP-YMCNE--CGKAFSVYSSL 459

Qy        187 QFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNEL 244

Db        460 TTHQVIHTGEKPYKCNECGKVFTQNSHLARHRGIH---TGEKPYKCNEC--GKVFRHNSY 514

Qy        245 VEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLD 304

Db        515 LSRHQRIHTGEKP---YKYNEYGK---AFSEHSNLTTHQVIHTGEKPYKCNE-------- 560

Qy        305 EKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRR 364

Db        561 ---------------CGKVFTQNSHLAR--HRRVHTGGKPYQC--NECGKAFSQTSKLAR 601

Qy        365 HLKWH 369

Db        602 HQRVH 606
```

Search completed: February 8, 2005, 15:34:42
Job time : 131 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:14:32 ; Search time 41 Seconds
(without alignments)
966.861 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	837.5	37.6	429	2	S20050	transcription fact
2	495	22.2	374	2	T37676	zinc-finger protei
3	400	17.9	803	2	S26823	zinc finger protei
4	396.5	17.8	595	2	G02075	transcription repr
5	395.5	17.7	363	2	I38937	DNA/RNA-binding pr
6	393.5	17.7	1191	2	S35305	zinc finger protei
7	375.5	16.8	1350	2	S00647	finger protein - A
8	374.5	16.8	710	2	I48668	zinc finger protei
9	373	16.7	542	2	A54661	zinc finger protei
10	373	16.7	651	2	B32891	finger protein 2,
11	372	16.7	393	2	JN0533	finger protein pML
12	364.5	16.4	338	2	G01496	transcription fact
13	364	16.3	428	2	A32891	finger protein 1,
14	363.5	16.3	615	2	S06546	finger protein (cl
15	359.5	16.1	427	2	A35659	krueppel-related p
16	354	15.9	693	2	I37570	zinc finger protei
17	353.5	15.9	591	2	S65088	finger protein XFO
18	351	15.7	474	2	I54338	zinc finger protei
19	348.5	15.6	439	2	S06556	finger protein (cl
20	348.5	15.6	707	2	S68858	finger protein - m
21	345.5	15.5	335	2	JC1441	transcription fact
22	345.5	15.5	594	2	T12488	hypothetical prote
23	344	15.4	337	2	S60520	finger protein ZNF
24	344	15.4	580	2	A37107	spermatogenesis pr
25	343	15.4	686	2	A34612	zinc finger protei
26	342	15.3	488	2	S47072	finger protein HZF
27	341	15.3	336	2	S06578	finger protein (cl
28	341	15.3	347	2	S00549	developmental cont
29	340	15.3	339	2	JC1442	transcription fact
30	339	15.2	411	2	S10245	finger protein, te
31	338	15.2	247	2	S47070	finger protein HZF
32	338	15.2	383	2	C32891	finger protein 9,
33	337.5	15.1	728	2	A48830	probable transcrip
34	337	15.1	469	2	I38600	zinc finger protei
35	337	15.1	595	2	JC7779	Krueppel-associate
36	336	15.1	553	2	S22954	finger protein zfp
37	335.5	15.1	428	2	S03677	finger protein (cl
38	334.5	15.0	280	2	S06572	finger protein (cl
39	332	14.9	536	2	S06548	finger protein (cl
40	332	14.9	701	2	T14757	hypothetical prote
41	331	14.8	794	2	S59069	Z13 protein - mous
42	330.5	14.8	732	2	S47073	finger protein HZF
43	329.5	14.8	496	2	T08674	probable finger pr
44	329	14.8	589	2	I38598	zinc finger protei
45	329	14.8	728	2	A54603	transcription fact

ALIGNMENTS

RESULT 1
S20050
transcription factor IIIA - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein P9677.9; protein YPR186c; TFC2 protein
C;Species: Saccharomyces cerevisiae
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: S20050; A44086; S58816
R;Archambault, J.; Milne, C.A.; Schappert, K.T.; Baum, B.; Friesen, J.D.; Segall, J.
J. Biol. Chem. 267, 3282-3288, 1992
A;Title: The deduced sequence of the transcription factor TFIIIA from Saccharomyces cer
A;Reference number: S20050; MUID:92147684; PMID:1737784
A;Accession: S20050
A;Molecule type: DNA
A;Residues: 1-429 <ARC>
A;Cross-references: UNIPROT:P39933; EMBL:M80611; NID:g172902; PIDN:AAB08014.1; PID:g172
R;Woychik, N.A.; Young, R.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 3999-4003, 1992
A;Title: Genes encoding transcription factor IIIA and the RNA polymerase common subunit
A;Reference number: A44086; MUID:92237295; PMID:1570325
A;Accession: A44086
A;Molecule type: DNA
A;Residues: 1-429 <WOY>
A;Cross-references: GB:M90638
R;Miller, N.
submitted to the EMBL Data Library, April 1995
A;Description: The sequence of S. cerevisiae cosmid 9677.
A;Reference number: S58816
A;Accession: S58816
A;Molecule type: DNA
A;Residues: 1-429 <MIL>
A;Cross-references: EMBL:U25841; NID:g786295; PID:g786305; MIPS:YPR186c
C;Genetics:
A;Gene: SGD:PZF1; TFIIIA; TFC2
A;Cross-references: SGD:S0006390; MIPS:YPR186c
A;Map position: 16R
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; nucleus; transcription factor; zinc finger

Query Match 37.6%; Score 837.5; DB 2; Length 429;
Best Local Similarity 43.7%; Pred. No. 1.6e-43;
Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;

```
Qy          1 MSESDETKSISSLIS--SSSSSRPKKYICTYEGCDKAYNRPSLLEQH-LRTHSNDRPYKC 57
              :| |: ::|::|| |  ||||:||| | | |:|||||: |||:| :| |  |    | ::|
Db         23 ISRSESSESLNSLTSTRSSSSNRPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQC 82

Qy         58 TVDDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFEN 117
                | | |:| :||||| |: :||: ||| || ||||| :|| |||||:|||||  |  |
Db         83 --DKCAKSFVKKSHLERHLYTHSDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEG 140

Qy        118 CQEAFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGG--SPAYQCDHPG 175
              |   ||||  || ||||||   |||  ||| | || :|  |  |||       |||   |
```

```
Db        141 CNLRFYKHPQLRAHILSVHLHKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAG 200

Qy        176 CFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV 235

Db        201 CCKEFRIWSQLQSHIKNDHPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPD 260

Qy        236 GKFAKKNELVEHYNIFH-DGNIPDDL-LKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293

Db        261 MSFSRKHDLLTHYGSIHTEEDIPLELKYKISDIQQL--VQDHGVQLGN------SKHSNE 312

Qy        294 EDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLE-GSKSVSKLISNSGKKINCPKNNC 352

Db        313 QDEEKISNRLRKRRKLTENNNVEFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNC 372

Qy        353 DRMFSREYDLRRHL---KWHDDNLQRIESFLNSIEKEE 387

Db        373 SRTFKTKEKYEKHIDKHKVHELKLKILQ------EKEE 404


RESULT 2
T37676
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T37676
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21736
A;Accession: T37676
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-374 <MCD>
A;Cross-references: UNIPROT:Q9UTL5; EMBL:AL132675; PIDN:CAB59689.1; GSPDB:GN00066; SPDB:
A;Experimental source: strain 972h-; cosmid c144
C;Genetics:
A;Gene: SPDB:SPAC144.09c
A;Map position: 1
A;Introns: 50/3
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

  Query Match             22.2%;  Score 495;  DB 2;  Length 374;
  Best Local Similarity   32.8%;  Pred. No. 6.7e-23;
  Matches  123;  Conservative   58;  Mismatches  160;  Indels   34;  Gaps   13;

Qy         15 SSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLET 74

Db         13 SKNLRSAKKIFHCPYEECGKKYSRPSLLEQHLRTHSNERPFVCDYTGCSKAFYRKSHLKI 72

Qy         75 HIVSHSEKKPFHCSV--CGKGVNSRQHLKRHEITH--TKSFKCTFENCQEAFYKHQSLRH 130

Db         73 HKRCHTNVKPFSCHYDGCDAQFYTQQHLERHIEVHRKPKPYACTWEGCDECFSKHQQLRS 132

Qy        131 HILSVHEKTL----TCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC--FKNFQTWS 184

Db        133 HISACHTHLLPYPCTYQDCELRFATKQKLQNHVNRAHEKIISYSCPHESCVGHEGFEKWS 192

Qy        185 VLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV----GKFAK 240

Db        193 QLQNHIREAHVP-SCSICGRQFKTAAHLRHHVVLH--QTTLEERKTYHCPMEGCKKSFTR 249

Qy        241 KNELVEHYNIFHDGNIP---DDLLKETEVK-KLENLLDQGSKLNNLHELETEKLKVEEDE 296

Db        250 SSALKKHISVIHEGNMAFHCDSCGTKFGYKHMLQRHLERGT-CKKAHKPYINECGIKHDG 308

Qy        297 EDEEDSLDEKRSDVRSDSMS--AQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDR 354

Db        309 IEGVAIHDQKEKELSSNLVSDVAKKIINEVTG--HGYK--------EAREYSCSFPECNY 358

Qy        355 MFSREYDLRRHLKWH 369

Db        359 RFKRLYDMHRHLNSH 373


RESULT 3
S26823
zinc finger protein ZNF43 - human
N;Alternate names: zinc finger protein kox27
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S26823; I37967; S10416
R;Lovering, R.; Trowsdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A;Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell li
A;Reference number: S26823; MUID:91279444; PMID:1711675
A;Accession: S26823
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-803 <LOV>
A;Cross-references: UNIPROT:P17038; EMBL:X59244; NID:g38031; PIDN:CAA41932.1; PID:g38032
R;Thiesen, H.J.
New Biol. 2, 363-374, 1990
A;Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A;Reference number: I37949; MUID:91145339; PMID:2288909
A;Accession: I37967
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 476-531 <THI>
A;Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090
C;Genetics:
A;Gene: GDB:ZNF43; HTF6
A;Cross-references: GDB:128653
A;Map position: 19p13.1-19p12
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

  Query Match             17.9%;  Score 400;  DB 2;  Length 803;
  Best Local Similarity   31.9%;  Pred. No. 8.4e-17;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59

Db        367 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 422

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117

Db        423 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 480

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176

Db        481 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 537

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234

Db        538 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 593

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293

Db        594 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 647

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350

Db        648 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 704

Qy        351 NCDRMFSREYDLRRHLKWH 369

Db        705 -CGKAFNRPSNLIEHKKIH 722


RESULT 4
G02075
transcription repressor zinc finger protein 85 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C;Accession: G02075
R;Poncelet, D.A.
submitted to the EMBL Data Library, September 1995
```

```
A;Reference number: G09169
A;Accession: G02075
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-595 <PON>
A;Cross-references: UNIPROT:Q03923; EMBL:U35376; NID:g1017721; PIDN:AAA79179.1; PID:g101
C;Genetics:
A;Gene: GDB:ZNF85
A;Cross-references: GDB:132279
A;Map position: 19p12-19p12
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

  Query Match             17.8%;  Score 396.5;  DB 2;  Length 595;
  Best Local Similarity   31.7%;  Pred. No. 9.9e-17;
  Matches  120;  Conservative   48;  Mismatches  160;  Indels   51;  Gaps   17;

Qy         10 ISSLISSSS-SSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFR 68
              || |    |   :|   | |  | | ||:|  | | :| | |: ::||||  ::| ||| :
Db        186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTKHKRIHTGEKPYKC--EECGKAFNQ 241

Qy         69 KSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQ 126
               |:|  |   |: :||: |  |||  |     |  |:| ||  | :||  : | :|| :
Db        242 SSNLIKHKKIHTGEKPYKCEECGKAFNRFSTLTTHKIIHTGEKPYKC--KECGKAFNRSS 299

Qy        127 SL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSV 185
              :|  |  :   ||   |::| | | : | |  ||: | |  | |:|    | | |   :
Db        300 TLTTHRKIHTGEKPYKCEECGKAFKQSSNLTTHKIIHTGEKP-YKCKK--CGKAFNQSAH 356

Qy        186 LQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNE 243
              |  |  |        || ||||       |::| : |   |  | : |  |  ||  | :
Db        357 LTTHEVIHTGEKPYKCEKCGKAFNHFSHLTTHKIIH---TGEKPYKCKEC--GKAFKHSS 411

Qy        244 LVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSL 303
              :  : | | |       |  : |: |   :| |||       ||  |:   |:  |
Db        412 TLTKHKIIHTGE------KPYKSKECEKAFNQSSKL-------TEHKKIHTGEKPYE--- 455

Qy        304 DEKRSDVRSDSMSAQRSIKSFTASL--------EGSK-----SVSKLISNSGKKINCPKN 350
              ||     : | :  |  || |           :| |     :: |:|     |   |
Db        456 CEKCGKAFNQSSNLTRHKKSHTEEKPYKCEECGKGFKWPSTLTIHKIIHTGEKPYKC--E 513

Qy        351 NCDRMFSREYDLRRHLKWH 369
              | : |::   | :| | |
Db        514 ECGKAFNQSSKLTKHKKIH 532


RESULT 5
I38937
DNA/RNA-binding protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I38937
R;Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.
Gene 159, 215-218, 1995
A;Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA.
A;Reference number: I38937; MUID:95347600; PMID:7622052
A;Accession: I38937
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-363 <RES>
A;Cross-references: EMBL:U20272; NID:g644870; PIDN:AAA75623.1; PID:g644871
C;Superfamily: transcription factor IIIA

  Query Match             17.7%;  Score 395.5;  DB 2;  Length 363;
  Best Local Similarity   31.1%;  Pred. No. 6.7e-17;
  Matches   99;  Conservative   54;  Mismatches   96;  Indels   69;  Gaps   16;

Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
              :|:|||      |::  ||         |:::||::   |    |::   |: ||  |: :|
Db          7 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 66

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
              |: |  : | ||| |  ||  ||::|: :||| |:    | :  |:: :||:| |  |
Db         67 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 126

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
              | : |:||:|:: | ||| |: |      :|    | |  | | |  |||| :|   | |
Db        127 QKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 186

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
                | |   ||    :||: |  |::::| :
Db        187 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 241

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
                 :||:  ||:        | ||:|| |::|   : : |::   || || |  |  |  :
Db        242 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 297

Qy        251 FHDGNIPDDLLKETEVKK 268
              ||   ||    : :|||
Db        298 VHD---PDKKKMKLKVKK 312


RESULT 6
S35305
zinc finger protein ZNF91 - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 09-Jul-2004
C;Accession: S35305
R;Bellefroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; Pon
EMBO J. 12, 1363-1374, 1993
A;Title: Clustered organization of homologous KRAB zinc-finger genes with enhanced expr
A;Reference number: S35305; MUID:93223677; PMID:8467795
A;Accession: S35305
A;Molecule type: mRNA
A;Residues: 1-1191 <BEL>
A;Cross-references: UNIPROT:Q05481; EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186
A;Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residue
C;Genetics:
A;Gene: GDB:ZNF91; HPF7; HTF10
A;Cross-references: GDB:132284
A;Map position: 19p12-19p12
C;Keywords: DNA binding; zinc finger

  Query Match             17.7%;  Score 393.5;  DB 2;  Length 1191;
  Best Local Similarity   35.8%;  Pred. No. 3.2e-16;
  Matches   98;  Conservative   40;  Mismatches  109;  Indels   27;  Gaps   12;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              :| | | |  | | ||:::|| |  | | |: ::||||  ::| ||| : | | || : |
Db        905 TREKPYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 960

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : :||: |  |||       |  |:| ||  | :||  | | :|| :  :| ||  :
Db        961 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1018

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
              ||   |::| | | | |||  ||: | |  | |:|:   | | | : | |  | |: | :
Db       1019 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1074

Qy        197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253
                 || :|||       |: |   |   |  | : |  |  ||  :  : | |
Db       1075 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1129

Qy        254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287
              |       |  : :|     :| | | |  :: |
Db       1130 GE------KPYKCEKCCKAFNQSSILTNHKKIHT 1157


RESULT 7
S00647
finger protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C;Accession: S00647
R;Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A;Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
```

```
A;Reference number: S00647; MUID:88082679; PMID:2826129
A;Accession: S00647
A;Molecule type: mRNA
A;Residues: 1-1350 <RUI>
A;Cross-references: EMBL:X06021
A;Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiatc
C;Genetics:
A;Gene: fin
C;Keywords: DNA binding; zinc finger

  Query Match             16.8%;  Score 375.5;  DB 2;  Length 1350;
  Best Local Similarity   27.3%;  Pred. No. 4.5e-15;
  Matches  119;  Conservative   65;  Mismatches  171;  Indels   81;  Gaps   20;

Qy          11 SSLISSSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
               | |     : || | |:|::  |:| : : | | :| |||: :|||:|   :| | | :||
Db         340 SFLNHQQTHSREKPYLCSH--CNKGFIQNSDLVKHFRTHTGERPYQCA--ECHKGFIQKS 395

Qy          71 HLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL 128
                 |  |: :|: :||| || | |     |  | :|: |||  | :||:  :| : | :  :|
Db         396 DLVKHLRTHTGEKPFKCSHCDKKFTERSALAKHQRTHTGEKPYKCS--DCGKEFTQRSNL 453

Qy         129 -RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKH--------HGGSPAYQCDHPGCFKN 179
                |  :   |:   |  |:: | : | | :|:  |         |  :  ::|    |
Db         454 ILHQRIHTGERPYKCTLCDRTFIQNSDLVKHQKVHANLPLSDPHTANSPHKCSK--CDLT 511

Qy         180 FQTWSVLQFHIK--QSHPKLKCPKCGKGCVGKKGLSSHMLSH----------------DD 221
               |  ||    | |       | :| :| ||    |  | |:  |                :
Db         512 FSHWSTFMKHSKLHSGEKKFQCAECKKGFTQKSDLVKHIRVHTGEKPFKCLLCKKSFSQN 571

Qy         222 STMIKIW---------TCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENL 272
               | : | |              |  ||   | ::: |::|:   | |  |       :    :
Db         572 SDLHKHWRIHTGEKPFPCYTCD-KSFTERSALIKHHRT-HTGERPH------KCSVCQKG 623

Qy         273 LDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQR------------S 320
                 | | |        |:  :    |:    :   |      || :  ||
Db         624 FIQKSAL-------TKHSRTHTGEKPYPCTQCGKSFIQNSDLVKHQRIHTGEKPYHCTEC 676

Qy         321 IKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWHD-DNLQRIES 378
               | ||   |||  |    ::||:|   ||:  |:: | :  || :||  |: :|     :
Db         677 NKRFT---EGSSLVKHRRTHSGEKPYRCPQ--CEKTFIQSSDLVKHLVVHNGENPPAATA 731

Qy         379 FLN-SIEKEETPEGEP 393
               |   | :|    ||
Db         732 FHEILIRRENLTRSEP 747


RESULT 8
I48668
zinc finger protein 51 - mouse
N;Alternate names: finger protein zfec12
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I48668; A40984
R;Burke, P.S.; Don, J.; Wolgemuth, D.J.
Mamm. Genome 5, 387-389, 1994
A;Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of c
A;Reference number: I48668; MUID:94319090; PMID:8043957
A;Accession: I48668
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-710 <RES>
A;Cross-references: UNIPROT:Q61898; EMBL:X74855; NID:g488832; PIDN:CAA52847.1; PID:g4888
R;Crossley, P.H.; Little, P.F.R.
Proc. Natl. Acad. Sci. U.S.A. 88, 7923-7927, 1991
A;Title: A cluster of related zinc finger protein genes is deleted in the mouse embryoni
A;Reference number: A40984; MUID:91376058; PMID:1680234
A;Accession: A40984
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 672-710 <CRO>
A;Cross-references: GB:M74235
C;Genetics:
A;Gene: Zfp-51
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

  Query Match             16.8%;  Score 374.5;  DB 2;  Length 710;
  Best Local Similarity   29.6%;  Pred. No. 2.6e-15;
  Matches  104;  Conservative   54;  Mismatches  140;  Indels   53;  Gaps   15;

Qy          25 YICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEKKP 84
               | |  : | |::   | |::| | |: ::|||| :  |||:|   : |:||   |: :||
Db         311 YEC--KECGKSFIELSHLKRHYRIHTGEKPYKCEI--CDKSFTTTTTLKTHQKIHTGEKP 366

Qy          85 FHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILS-VHEKTLT 141
               : |  | |       ||:||:  ||  : ::|  : | ::|::  :|| |  |   |||
Db         367 YKCRECDKSFIHSSHLRRHQNVHTGERPYRC--KECDKSFHESATLREHEKSHTGEKTYK 424

Qy         142 CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKLKC 199
               |::|:| ||: : |  |  : | |    |:|    | |:| | | |: |  |        ||
Db         425 CRECDKSFTQRAYLRNHHNRVHTGERPYECKE--CGKSFTTCSTLRIHQTIHTGKKPYKC 482

Qy         200 PKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIPDD 259
               :|||        | :|    |      | : |  ||    |   : |  | :| | | |
Db         483 IECGKSFTNNSYLRTHQKVHSGE---KPYRCKECD-KSFTSCSTLKAHQSI-HTGEKPYK 537

Qy         260 LLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQR 319
               :      : :   | | |     : | :      | |          :| :||  :   :
Db         538 CI------ECDKSFTQVSHLRTHQRVHTGERPFRCTECD--------KSFIRSSHLREHQ 583

Qy         320 SIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWH 369
               |                    :||:|   |   :||  ||:  :|||| | |
Db         584 KI-----------------HSGEKPYKC--RDCDISFSQISNLRRHQKLH 614


RESULT 9
A54661
zinc finger protein ZNF41 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 09-Jul-2004
C;Accession: A54661; I54227
R;Franze, A.; Archidiacono, N.; Rocchi, M.; Marino, M.; Grimaldi, G.
Genomics 9, 728-736, 1991
A;Title: Isolation and expression analysis of a human zinc finger gene (ZNF41) located o
A;Reference number: A54661; MUID:91244317; PMID:2037297
A;Accession: A54661
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-542 <FRA>
A;Cross-references: UNIPROT:P51814; GB:M36548; GB:M92443; NID:g340443; PIDN:AAA61312.1;
C;Genetics:
A;Gene: GDB:ZNF41
A;Cross-references: GDB:125865; OMIM:314995
A;Map position: Xp11.3-Xp11.23
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

  Query Match             16.7%;  Score 373;  DB 2;  Length 542;
  Best Local Similarity   29.0%;  Pred. No. 2.4e-15;
  Matches  115;  Conservative   59;  Mismatches  156;  Indels   66;  Gaps   21;

Qy          14 ISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLE 73
               |   : :  | | |    | ||: | | |  | | |: ::|| |   || ||| :|||
Db         163 IHQKTHTGEKHYEC--NECGKAFTRKSALRMHQRIHTGEKPYVCA--DCGKAFIQKSHFN 218

Qy          74 THIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH 131
               ||   |: :||: || |||    :  |  |:  ||  | : ||    | : |    :
Db         219 THQRIHTGEKPYECSDCGKSFTKKSQLHVHQRIHTGEKPYICT--ECGKVFTHRTN---- 272

Qy         132 ILSVHEKTLT------CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSV 185
                |: |:|| |       | :| | ||  | | :|:  | |  | |:|:  || | |   |
Db         273 -LTTHQKTHTGEKPYMCAECGKAFTDQSNLIKHQKTHTGEKP-YKCN--GCGKAFIWKSR 328

Qy         186 LQFHIKQSH---PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKK 241
               |: | ::||      :|  |||  : |  || |   |  | : |  | || | :|
Db         329 LKIH-QKSHIGERHYECKDCGKAFIQKSTLSVHQRIH---TGEKPYVCPEC--GKAFIQK 382
```

```
Qy        242 NELVEHYNIFHDGNIPDD-------LLKETEVKKLENLLDQGSKLNNLHE---------- 284
              :  : |: | | |  | :         |:::: ::   :  | | |   |
Db        383 SHFIAHHRI-HTGEKPYECSDCGKCFTKKSQL-RVHQKIHTGEKPNICAECGKAFTDRSN 440

Qy        285 -LETEKLKVEEDEEDEEDS----LDEKRSDVRSDSMSAQRSI------KSFTASLEGSKS 333
               :  :|:   |   :  |        : | ::   | : :|        |:|    : : |
Db        441 LITHQKIHTREKPYECGDCGKTFTWKSRLNIHQKSHTGERHYECSKCGKAFIQ--KATLS 498

Qy        334 VSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
              : ::|    |   |    | : |:   :| :| | |
Db        499 MHQIIHTGKKPYAC--TECQKAFTDRSNLIKHQKMH 532


RESULT 10
B32891
finger protein 2, placental - human
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C;Accession: B32891
R;Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A
DNA 8, 377-387, 1989
A;Title: The human genome contains hundreds of genes coding for finger proteins of the K
A;Reference number: A32891; MUID:89377476; PMID:2505992
A;Accession: B32891
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-651 <BEL>
A;Cross-references: GB:M27878
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: tandem repeat

  Query Match             16.7%;  Score 373;  DB 2;  Length 651;
  Best Local Similarity   27.4%;  Pred. No. 2.9e-15;
  Matches  128;  Conservative   57;  Mismatches  146;  Indels  136;  Gaps   24;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH--- 79
              | | |:   | ||::: | | | |||: ::||:|   :| ||| ||||| :|  :|
Db        174 KPYNCSQ--CGKAFSQKSQLTSHQRTHTGEKPYEC--GECGKAFSRKSHLISHWRTHTGE 229

Qy         80 ------------SEK------------KPFHCSVCGKGVNSRQHLKRHEITH--TKSFK 112
                          |||            ||| |  |||  : :  |  |  ||  || |
Db        230 KPYGCNECGRAFSEKSNLINHQRIHTGEKPFECRECGKAFSRKSQLVTHHRTHTGTKPFG 289

Qy        113 CTFENCQEAFY-KHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSP---- 167
              |:  :|::||: | : :||  :   ||   | :| | |   | |  |:  | |  |
Db        290 CS--DCRKAFFEKSELIRHQTIHTGEKPYECSECRKAFRERSSLINHQRTHTGEKPHGCI 347

Qy        168 ------------AYQCDHPG--------CFKNFQTWSVLQFHIKQSHPKLK---CPKCG 203
                         ::|  | |         | | |   | | |  :::|   |    | :||
Db        348 QCGKAFSQKSHLISHQMTHTGEKPFICSKCGKAFSRKSQLVRH-QRTHTGEKPYECSECG 406

Qy        204 KGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDLL- 261
              |   |  |::|   |   | | : |  |  || | :|: |: |   |  | |  | :
Db        407 KAFSEKLSLTNHQRIH---TGEKPYVCSEC--GKAFCQKSHLISHQRT-HTGEKPYECSE 460

Qy        262 --------------------KETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
                                 | | :  |    | |:||  |:      ::   |:  |
Db        461 CGKAFGEKSSLATHQRTHTGEKPYECRDCEKAFSQKSQLNT-HQ------RIHTGEKPYE 513

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSRE 359
               ||  |                |  | |: :  | :::|:|   |  | | : |  :
Db        514 CSLCRK-----------------AFFEKSELIRHLRTHTGEKPYEC--NECRKAFREK 552

Qy        360 YDLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNE 406
              |  |        ||| :     |  |   |:   :|:   |:|::
Db        553 SSLINH--------QRIHTGEKPFECSEC--GKAFSRKSH--LIPHQ 587


RESULT 11
JN0533
finger protein pMLZ-4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: JN0533
R;Brady, J.P.; Piatigorsky, J.
Gene 124, 207-214, 1993
A;Title: Cloning and characterization of a novel zinc-finger protein encoding cDNA from
A;Reference number: JN0533; MUID:93185925; PMID:8444344
A;Accession: JN0533
A;Molecule type: mRNA
A;Residues: 1-393 <BRA>
A;Cross-references: UNIPROT:Q03309; GB:M98502; NID:g200406; PIDN:AAA39949.1; PID:g20040
A;Experimental source: eye lens
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; eye lens; zinc finger
F;1-62/Region: acidic
F;63-391/Region: zinc finger

  Query Match             16.7%;  Score 372;  DB 2;  Length 393;
  Best Local Similarity   29.4%;  Pred. No. 1.9e-15;
  Matches  110;  Conservative   57;  Mismatches  163;  Indels   44;  Gaps   17;

Qy         13 LISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHL 72
              |:   : :  : :||:   | ||::: | | :| :||: ||||||   :| | | | |||
Db         49 LVHKEAHAGIRYHICSQ--CGKAFSQISDLNRHQKTHTGDRPYKCY--ECGKGFSRSSHL 104

Qy         73 ETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-R 129
                 |  :|: ::|: |: |||       || :|:  ||  |  |||   | :|      | :
Db        105 IQHQRTHTGERPYDCNECGKSFGRSSHLIQHQTIHTGEKPHKCT--ECAKASAASPHLIQ 162

Qy        130 HHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWS--VLQ 187
              |      ||   |::| | |:| | ||||:  | |  | |:|    | : |  |  :
Db        163 HQRTHSGEKPYECEECGKSFSRSSHLAQHQRTHTGEKP-YECHE--CGRGFSERSDLIKH 219

Qy        188 FHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEH 247
              : :       || :|||       |  |  :|   |  | : |: |    |:: : ||:|
Db        220 YRVHTGERPYKCDECGKNFSQNSDLVRHRRAH---TGEKPYHCNECG-ENFSRISHLVQH 275

Qy        248 YNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----EDEEDS 302
                  | |        |  |          : | |    :: | :   | :|      :  |
Db        276 QRT-HTGE------KPYECTACGKSFSRSSHLITHQKIHTGEKPYECNECWRSFGERSDL 328

Qy        303 LDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLIS----NSGKK-INCPKNNCDRMFS 357
              :  :|:          :  | || |       | ||:    ::|:|   |    ||: ||
Db        329 IKHQRTHTGEKPYECVQCGKGFTQS-------SNLITHQRVHTGEKPYEC--TECDKSFS 379

Qy        358 REYDLRRHLKWHDD 371
              |   | :| : | |
Db        380 RSSALIKHKRVHTD 393


RESULT 12
G01496
transcription factor IIIA - human (fragment)
N;Alternate names: TFIIIA
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C;Accession: G01496
R;Becker, K.G.
submitted to the EMBL Data Library, August 1994
A;Reference number: G07520
A;Accession: G01496
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-338 <BEC>
A;Cross-references: EMBL:U14134; NID:g551534; PIDN:AAA21873.1; PID:g551535
C;Genetics:
A;Gene: GDB:GTF3A; TFIIIA
A;Cross-references: GDB:434744; OMIM:600860
A;Map position: 13q12.3-13q13.1
C;Superfamily: transcription factor IIIA

  Query Match             16.4%;  Score 364.5;  DB 2;  Length 338;
  Best Local Similarity   29.0%;  Pred. No. 4.6e-15;
  Matches   89;  Conservative   49;  Mismatches  102;  Indels   67;  Gaps   11;
```

```
Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
              :|:|||       |::  ||         |:::||::  |   |::   |: ||  |: :|
Db          7 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 66

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
              |: |  : | ||| |  ||  ||::|: :||| |:   | :  |:: :||:| |  |
Db         67 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 126

Qy        108 TKSFKCTFENCQEAFYKHQSLR------------------------------HHILSVH 136
               | : |:||:|:: | ||| |:                              |:   |
Db        127 QKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKLCCQKGCSFVAKTWTELLKHVRETH 186

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
              :: : |: | | | |  | ||  |      :|  || : : |  || ||  | :
Db        187 KEEILCEVCRKTFKRKDYLKQHMKTHAPERDVCRCPREGCGRTYTTVFNLQSHILSFHEE 246

Qy        197 LK---CPK--CGKGCVGKKGLSSHMLSHD-DSTMIKIWTCDYCDVGKFAKKNELVEHYNI 250
              :   |    |||    |: |: | : || |   :|:       | |  :|  |  | :
Db        247 SRPFVCEHAGCGKTFAMKQSLTRHAVVHDPDKKKMKL------KVKKSREKRSLASHLS- 299

Qy        251 FHDGNIP 257
                | ||
Db        300 ---GYIP 303


RESULT 13
A32891
finger protein 1, placental - human
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 09-Jul-2004
C;Accession: A32891
R;Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A
DNA 8, 377-387, 1989
A;Title: The human genome contains hundreds of genes coding for finger proteins of the K
A;Reference number: A32891; MUID:89377476; PMID:2505992
A;Accession: A32891
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-428 <BEL>
A;Cross-references: UNIPROT:P51522; GB:M27877
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: tandem repeat

  Query Match             16.3%;  Score 364;  DB 2;  Length 428;
  Best Local Similarity   29.1%;  Pred. No. 6.4e-15;
  Matches  116;  Conservative   50;  Mismatches  155;  Indels   78;  Gaps   18;

Qy         32 CDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSVCG 91
              | | :|: | |  | | |: ::||||  ::| | |   |||  |   |: :||: |: ||
Db         38 CGKIFNKKSNLASHQRIHTGEKPYKC--NECGKVFHNMSHLAQHRRIHTGEKPYKCNECG 95

Qy         92 KGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHEKTLTCKQCNKV 148
              |  |   || :|:  ||  | :||    | : |::   | :|  :   ||   | :| ||
Db         96 KVFNQISHLAQHQRIHTGEKPYKC--NECGKVFHQISHLAQHRTIHTGEKPYECNKCGKV 153

Qy        149 FTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKLKCPKCGKGC 206
              |:| | | || : | |  | |:|:   | | |  | |  |  |       || :|||
Db        154 FSRNSYLVQHLIIHTGEKP-YRCN--VCGKVFHHISHLAQHQRIHTGEKPYKCNECGKVF 210

Qy        207 VGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIP-------- 257
              |  | :|   |   |  | : |: |  || |: |: || |: | | |  |
Db        211 SHKSSLVNHWRIH---TGEKPYKCNEC--GKVFSHKSSLVNHWRI-HTGEKPYKCNECGK 264

Qy        258 ---------DDLL-----KETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSL 303
                      |:     |  :  : :    | | |   | : | :   : ||  :  |
Db        265 VFSRNSYLAQHLIIHAGEKPYKCDECDKAFSQNSHLVQHHRIHTGEKPYKCDECGKVFSQ 324

Qy        304 D------------EKRSDVR--------SDSMSAQRSI-------------KSFTASLEG 330
              :             ||            : |::  | |             |:|  |:
Db        325 NSYLAYHWRIHTGEKAYKCNECGKVFGLNSSLAHHRKIHTGEKPFKCNECGKAF--SMRS 382

Qy        331 SKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                | :    |    |   |  | | ::|     | || ::|
Db        383 SLTNHHAIHTGEKHFKC--NECGKLFRDNSYLVRHQRFH 419


RESULT 14
S06546
finger protein (clone XlcOF7.1) - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
C;Accession: S06546
R;Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poeting, A.; Knoech
J. Mol. Biol. 208, 639-659, 1989
A;Title: Second-order repeats in Xenopus laevis finger proteins.
A;Reference number: S05632; MUID:90040698; PMID:2509712
A;Accession: S06546
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-615 <NIE>
A;Cross-references: UNIPROT:P18751
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

  Query Match             16.3%;  Score 363.5;  DB 2;  Length 615;
  Best Local Similarity   27.6%;  Pred. No. 1e-14;
  Matches  119;  Conservative   48;  Mismatches  165;  Indels   99;  Gaps   18;

Qy         11 SSLISSSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
              |||:    :   :|  |  | ||| :   | |  | |||: ::|: |:  :| | |   |
Db        128 SSLVHHQRTHIGEKPFCCSE-CDKCFASSSELNIHQRTHTGEKPFSCS--ECGKCFTNHS 184

Qy         71 HLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL 128
              |   | : |: :||| || |||    |   |  |  |||  |:| |:    | :  |  |
Db        185 HFAHHQMIHTGEKPFCCSKCGKCFASSSDLTFHRRTHTREKTFSCS--ECGKCFSNHSHL 242

Qy        129 ------------------------------------------------------------RH 130
                                                                           ||
Db        243 ARHQMIHTGEKPFCCSECGKCFSSSSGLTAHQQRTHMKVKPFSCSACGKCFSNRSHLIRH 302

Qy        131 HILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHI 190
               ::   ||  :| :| | |: || ||:|:: | |  | : |    | | | : | | ||
Db        303 QMIHTGEKPFSCFECRKCFSNPSNLARHQMTHTGEKP-FSCSE--CGKCFASSSDLTFH- 358

Qy        191 KQSHPKLK---CPKCGKGCVGKK-GLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVE 246
              ::|   |   | :||| |  ||  |  | :|          || |    ||  :||
Db        359 HRTHTGEKPFSCSECGK-CYSKKSSLVHHQRTHTGEKPFSCSKCDKC----FASSSEL-- 411

Qy        247 HYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEK 306
              ||    :  :     :|  |      | |:   :|  | : :   | ||    |
Db        412 --NIHQRTHTGEKAFSCSECGKCFTNRSQLSRHQMIHTGE-KPISCPECEECFVSS---- 464

Qy        307 RSDVRSDSMSAQRSIKSFTASLEGSKSVS--------KLISNSGKKINCPKNNCDRMFSR 358
              | : :    | | :| |:  ||  |  |         ::|    | :|  : | : || |
Db        465 -SQLTAHQQQAHRMVKPFSC-LECGKCFSNRSNFARHQMIHTGEKPFSC--SECRKGFSN 520

Qy        359 EYDLRRHLKWH 369
              :  | ||  |
Db        521 QSSLARHQMTH 531


RESULT 15
A35659
krueppel-related protein H-plk - human
C;Species: Homo sapiens (man)
C;Date: 28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 09-Jul-2004
C;Accession: A35659
R;Kato, N.; Shimotohno, K.; VanLeeuwen, D.; Cohen, M.
Mol. Cell. Biol. 10, 4401-4405, 1990
A;Title: Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger pro
A;Reference number: A35659; MUID:90318410; PMID:2115127
A;Accession: A35659
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <KAT>
```

```
A;Cross-references: UNIPROT:Q02313; GB:M55422; NID:g184342; PIDN:AAA36010.1; PID:g184343
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

  Query Match            16.1%;  Score 359.5;  DB 2;  Length 427;
  Best Local Similarity  32.0%;  Pred. No. 1.2e-14;
  Matches   99;  Conservative   36;  Mismatches  121;  Indels   53;  Gaps   13;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              :|    | |   |    :|:|   | | :| | |: ::||||    :| ||| : |||  |    |
Db        132 TRVNFYKC--EAYGRAFNWSSTLNKHKRIHTGEKPYKC--KECGKAFNQTSHLIRHKRIH 187

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHTKS--FKCTFENCQEAFYKHQSL-RHHILSVH 136
              :|:||: |  |||  |    |  | | ||    :||  | |  || :   |  | ::
Db        188 TEEKPYKCEECGKAFNQSSTLTTHNIIHTGEIPYKC--EKCVRAFNQASKLTEHKLIHTG 245

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKH-------------------------HGGSPAY 169
              ||   |::| | | | ||| :||  |                         | |   |
Db        246 EKRYECEECGKAFNRSSKLTEHKYIHTGEKLYKCEECGKAFNQSSTLTTHKRIHSGEKPY 305

Qy        170 QCDHPGCFKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKI 227
              :|:   | | |: :| |  |  |        || :|||       |: | : |   |  |
Db        306 KCEE--CGKAFKQFSNLTDHKKIHTGEKPYKCEECGKAFNQLSNLTRHKVIH---TGEKP 360

Qy        228 WTCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287
              : |  |  ||   ::  :  : | | |  |    :  :|  |       |||:   :: |
Db        361 YKCGEC--GKAFNQSSALNTHKIIHTGENPHKCRESGKVFHL------SSKLSTCKKIHT 412

Qy        288 -EKL-KVEE 294
               ||| | ||
Db        413 GEKLYKCEE 421


Search completed: February  8, 2005, 15:22:57
Job time : 51 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 8, 2005, 15:11:59 ; Search time 179 Seconds
(without alignments)
1178.640 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1257	56.4	452	2	Q6BUY9	Q6buy9 debaryomyce
2	850	38.1	472	2	Q6CWK1	Q6cwk1 kluyveromyc
3	837.5	37.6	429	1	TF3A_YEAST	P39933 saccharomyc
4	819	36.7	476	2	Q758N3	Q758n3 ashbya goss
5	809.5	36.3	394	2	Q6FSJ8	Q6fsj8 candida gla
6	495	22.2	374	1	TF3A_SCHPO	Q9utl5 schizosacch
7	470.5	21.1	655	2	Q6C2U4	Q6c2u4 yarrowia li
8	411.5	18.5	634	2	Q6R5P3	Q6r5p3 mus musculu
9	410.5	18.4	634	2	Q6PEP4	Q6pep4 mus musculu
10	410.5	18.4	634	2	Q8BI85	Q8bi85 mus musculu
11	407	18.3	611	2	Q7RY67	Q7ry67 neurospora
12	400	17.9	404	2	Q8N7Q3	Q8n7q3 homo sapien
13	400	17.9	803	1	ZN43_HUMAN	P17038 homo sapien
14	398	17.9	574	1	Z492_HUMAN	Q9p255 homo sapien
15	397.5	17.8	395	2	O43693	O43693 homo sapien
16	397.5	17.8	423	1	TF3A_HUMAN	Q92664 homo sapien
17	396.5	17.8	595	1	ZN85_HUMAN	Q03923 homo sapien
18	395	17.7	616	1	ZN93_HUMAN	P35789 homo sapien
19	393.5	17.7	1191	1	ZN91_HUMAN	Q05481 homo sapien
20	392.5	17.6	542	2	Q8IYN0	Q8iyn0 homo sapien
21	392	17.6	555	2	Q96N38	Q96n38 homo sapien
22	392	17.6	644	2	Q8ND40	Q8nd40 homo sapien
23	392	17.6	661	2	Q86W65	Q86w65 homo sapien
24	391	17.5	592	2	Q8N8Q4	Q8n8q4 homo sapien
25	390	17.5	540	2	Q8N211	Q8n211 homo sapien
26	390	17.5	576	2	Q96N22	Q96n22 homo sapien
27	389	17.5	891	2	Q6ZNA1	Q6zna1 homo sapien
28	388.5	17.4	501	2	Q9VG72	Q9vg72 drosophila
29	388	17.4	568	2	Q8TD23	Q8td23 homo sapien
30	387.5	17.4	839	1	Z347_HUMAN	Q96se7 homo sapien
31	385.5	17.3	297	2	Q8BJ46	Q8bj46 mus musculu
32	385	17.3	576	1	Z431_HUMAN	Q8tf32 homo sapien
33	384	17.2	753	2	Q6P8L0	Q6p810 mus musculu
34	383	17.2	913	1	Z228_HUMAN	Q9uju3 homo sapien
35	382	17.1	511	2	Q8VEC1	Q8vec1 mus musculu
36	381	17.1	493	2	Q921H6	Q921h6 mus musculu
37	379.5	17.0	710	2	Q91VP4	Q91vp4 mus musculu
38	379	17.0	600	2	Q8CE00	Q8ce00 mus musculu
39	379	17.0	766	2	Q8C4J2	Q8c4j2 mus musculu
40	378.5	17.0	914	1	ZF62_MOUSE	Q8c827 mus musculu
41	378	17.0	393	2	Q923E2	Q923e2 mus musculu
42	378	17.0	452	2	Q8BPP0	Q8bpp0 mus musculu
43	378	17.0	452	2	Q8BQC2	Q8bqc2 mus musculu
44	378	17.0	461	2	Q69ZC6	Q69zc6 mus musculu
45	378	17.0	613	2	Q9H7R5	Q9h7r5 homo sapien

ALIGNMENTS

RESULT 1
Q6BUY9
ID Q6BUY9 PRELIMINARY; PRT; 452 AA.
AC Q6BUY9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to sp|P39933 Saccharomyces cerevisiae YPR186c TFC2
DE TFIIIA.
GN ORFNames=DEHA0C07623g;
OS Debaryomyces hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; CR382135; CAG86038.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR SMART; SM00355; ZnF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
SQ SEQUENCE 452 AA; 52049 MW; D8224533C04B0269 CRC64;

Query Match 56.4%; Score 1257; DB 2; Length 452;
Best Local Similarity 55.6%; Pred. No. 2.5e-64;
Matches 234; Conservative 62; Mismatches 93; Indels 32; Gaps 10;

```
Qy          3 ESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDC 62
              || || |:||  :||||:||| |||     | |||::||||||| |:|:|:||:||:   |
Db          2 ESSETASVSSR-ASSSSARPKNYICDEPDCGKAYSKPSLLEQHKRSHTNERPFKCSETGC 60

Qy         63 DKAFFRKSHLETHIVSH--SEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||:| |||||: |::||   | ||| |: ||||||: |||||||||||||| |||| | |
Db         61 DKSFLRKSHLQAHLLSHEDQESKPFQCATCGKGVNTLQHLKRHEITHTKSFICTFEGCNE 120

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              :|||||||||| |||||| | | :||| | || :|||| :|:|  |||||||| ||| ||
Db        121 SFYKHQSLRHHTLSVHEKKLLCNKCNKSFNRPYRLAQHNIKYHSDSPAYQCDHQGCFGNF 180

Qy        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240
              ||| || |||  |||:||| |||||||||| |||  ||:  ::|:| |:||::||| |
Db        181 MTWSALQLHIKTEHPKIKCPICGKGCVGKKGLRSHMNIHDEEKIVKLWNCNYCNIGKFVK 240

Qy        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSK----LNNLHELETEK-LKVEED 295
              | :|::||| :|| |:|||||  | ::|: || :       | |  |:::  ::|  |
Db        241 KADLIDHYNTYHDRNLPDDLLKPIEREQLDKLLSEKDSTNIDFNTLEGLQSKGFVEVPSD 300

Qy        296 EEDEEDSLDEKRSDVRSDSMSA-QRSIKSFTASLE-GSKSVSKLISNS--GKKINCPKNN 351
              || ||       :|  |: : |  :|: |   :|| |  |:  || |:   :|: ||| |
Db        301 EEKEE-------NDYMSNGLQASHKSLNSLNLTLESGKASIVDLILNNYLKRKVPCPKKN 353

Qy        352 CDRMFSREYDLRRHLKWHDDNLQRIESFLNSIEKEET------------PEG-EPLVKKA 398
              |||||||:||| ||||||: :|::||:|| |:| ||             ||  ||::
Db        354 CDRMFSRDYDLERHLKWHESHLKKIEAFLKSLELEEKVPDSNMHDKRLFPENDEPVMATK 413

Qy        399 R 399
              |
Db        414 R 414
```

RESULT 2
Q6CWK1
ID   Q6CWK1   PRELIMINARY;   PRT;   472 AA.
AC   Q6CWK1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similarities with sp|P39933 Saccharomyces cerevisiae YPR186c TFC2
DE   TFIIIA.
GN   ORFNames=KLLA0B03454g;
OS   Kluyveromyces lactis NRRL Y-1140.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX   NCBI_TaxID=284590;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL Y-1140;
RG   Genolevures;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL Y-1140;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382122; CAH02081.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 8.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
SQ   SEQUENCE   472 AA; 54229 MW; 5AE712DB26674A4D CRC64;

```
  Query Match             38.1%;  Score 850;  DB 2;  Length 472;
  Best Local Similarity   45.3%;  Pred. No. 5.6e-41;
  Matches  169;  Conservative   48;  Mismatches  128;  Indels   28;  Gaps    7;

Qy          9 SISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQH-LRTHSNDRPYKCTVDDCDKAFF 67
              |::| :| ||| | | | | |:||:|:: ||||| :| |  |   :|:||  | | | |
Db         90 SVTSTVSVSSSQRSKVYFCDYQGCNKSFTRPSLLTEHQLTVHHGIKPFKC--DTCGKEFA 147

Qy         68 RKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQS 127
              :||||  |: ||:: ||| ||:||||| :|| ||||||||||| |::| | |:||||
Db        148 KKSHLNRHMFSHTDDKPFTCSICGKGVTTRQQLKRHEITHTKSFHCSYEGCNESFYKHPQ 207

Qy        128 LRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHG--GSPAYQCDHPGCFKNFQTWSV 185
              || |||||||| ||| |||| | || :|  |  |||    :  ||||   |   | |||
Db        208 LRSHILSVHEKKLTCPHCNKTFQRPYRLKNHIDKHHNPESTGMYQCDFLSCTDVFSTWSS 267

Qy        186 LQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELV 245
              || |||| |||| || ||| || : || :||: ||:| : | | |   |    |||| :||
Db        268 LQQHIKQCHPKLPCPICGKPCVAESGLRNHMMIHDESLVTKNWKCSSCPDTSFAKKTQLV 327

Qy        246 EHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEED------- 298
              :||   |    |: ||  |:  :     |   |:|          :|: | |:
Db        328 QHYEESHKDVDPELLLLPTDADE-----DTNIKVNT-------DIKISETEDSFSSGKKT 375

Qy        299 --EEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMF 356
                :  |  :||       |: |:  || :  |  |  : | |::  |:|  ||  ||:| |
Db        376 LLDFPSFSKKRKVDELGSVEAEIKIKKYIES--GKTTKSLLLNTVGQKRKCPYTNCNRTF 433

Qy        357 SREYDLRRHLKWH 369
              :     |:: |
Db        434 KSKEKFDTHIEKH 446
```

RESULT 3
TF3A_YEAST
ID   TF3A_YEAST   STANDARD;   PRT;   429 AA.
AC   P39933;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Transcription factor IIIA (TFIIIA).
GN   Name=TFC2; Synonyms=PZF1, TFIIIA; OrderedLocusNames=YPR186C;
GN   ORFNames=P9677.9;
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92147684; PubMed=1737784;
RA   Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA   Segall J.;
RT   "The deduced sequence of the transcription factor TFIIIA from
RT   Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT   TFIIIA.";
RL   J. Biol. Chem. 267:3282-3288(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92237295; PubMed=1570325;
RA   Woychik N.A., Young R.A.;
RT   "Genes encoding transcription factor IIIA and the RNA polymerase
RT   common subunit RPB6 are divergently transcribed in Saccharomyces
RT   cerevisiae.";

```
RL   Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
CC   -!- FUNCTION: Interacts with the internal control region (ICR) of
CC       approximately 50 bases within the 5S RNA genes, is required for
CC       correct transcription of these genes by RNA polymerase III. Also
CC       binds the transcribed 5S RNA's.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80611; AAB08014.1; -.
DR   EMBL; M90638; -; NOT_ANNOTATED_CDS.
DR   EMBL; U25841; AAB64615.1; -.
DR   PIR; S20050; S20050.
DR   HSSP; P25490; 1UBD.
DR   GermOnline; 144451; -.
DR   TRANSFAC; T03530; -.
DR   SGD; S000006390; PZF1.
DR   GO; GO:0005667; C:transcription factor complex; IPI.
DR   GO; GO:0003709; F:RNA polymerase III transcription factor act. . .; IPI.
DR   GO; GO:0006384; P:transcription initiation from Pol III promoter; IPI.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 9.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   DNA-binding; Metal-binding; Nuclear protein; Repeat; RNA-binding;
KW   Transcription regulation; Zinc-finger.
FT   DOMAIN       24     43       Ser-rich.
FT   ZN_FING      49     74       C2H2-type.
FT   ZN_FING      80    102       C2H2-type.
FT   ZN_FING     108    130       C2H2-type.
FT   ZN_FING     134    159       C2H2-type.
FT   ZN_FING     163    186       C2H2-type.
FT   ZN_FING     194    219       C2H2-type.
FT   ZN_FING     222    244       C2H2-type.
FT   ZN_FING     253    277       C2H2-type.
FT   DOMAIN      321    327       Arg/Lys-rich (basic).
FT   ZN_FING     365    389       C2H2-type.
SQ   SEQUENCE   429 AA;  50027 MW;  209B1EDEA20422D9 CRC64;

  Query Match             37.6%;  Score 837.5;  DB 1;  Length 429;
  Best Local Similarity   43.7%;  Pred. No. 2.6e-40;
  Matches  174;  Conservative   62;  Mismatches  135;  Indels   27;  Gaps   11;

Qy          1 MSESDETKSISSLIS--SSSSSRPKKYICTYEGCDKAYNRPSLLEQH-LRTHSNDRPYKC 57
              :| |: ::|::|| |  ||||:||| | | |:|||||: |||:| :| |  |    | ::|
Db         23 ISRSESSESLNSLTSTRSSSSNRPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQC 82

Qy         58 TVDDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFEN 117
                | | |:| :||||| |: :||: ||| || ||||| :|| |||||:|||||| |  |
Db         83 --DKCAKSFVKKSHLERHLYTHSDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEG 140

Qy        118 CQEAFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGG--SPAYQCDHPG 175
              |   ||||  || ||||||   |||  ||| | || :|  |  |||        |||   |
Db        141 CNLRFYKHPQLRAHILSVHLHKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAG 200

Qy        176 CFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV 235
              | | |: || || |||  ||||||| | | |||: ||  ||: |||| : | | |  |
Db        201 CCKEFRIWSQLQSHIKNDHPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPD 260

Qy        236 GKFAKKNELVEHYNIFH-DGNIPDDL-LKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293
                |::|::|: ||   | : :|| :|  | :::::|  :  | | :| |        |   |
Db        261 MSFSRKHDLLTHYGSIHTEEDIPLELKYKISDIQQL--VQDHGVQLGN------SKHSNE 312

Qy        294 EDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLE-GSKSVSKLISNSGKKINCPKNNC 352
              :|||   : | ::|    ::::   ::       || |   :: |::  |:|  |  |||
Db        313 QDEEKISNRLRKRRKLTENNNVEFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNC 372

Qy        353 DRMFSREYDLRRHL---KWHDDNLQRIESFLNSIEKEE 387
              | |  :    :|:   | |:  |: ::      ||||
Db        373 SRTFKTKEKYEKHIDKHKVHELKLKILQ------EKEE 404


RESULT 4
Q758N3
ID   Q758N3      PRELIMINARY;      PRT;   476 AA.
AC   Q758N3;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   AEL278Wp.
GN   ORFNames=AEL278W;
OS   Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX   NCBI_TaxID=33169;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10895;
RA   Gates K., Dietrich F.S., Brachat S., Voegeli S.E., Lerch A.,
RA   Philippsen P., Gaffney T.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE016818; AAS52406.1; -.
DR   HSSP; P03001; 1UN6.
DR   AGD; AEL278W; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 8.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
SQ   SEQUENCE   476 AA;  53433 MW;  D1487545D4E6BE0D CRC64;

  Query Match             36.7%;  Score 819;  DB 2;  Length 476;
  Best Local Similarity   40.8%;  Pred. No. 3.4e-39;
  Matches  165;  Conservative   65;  Mismatches  136;  Indels   38;  Gaps    9;

Qy          2 SESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRT-HSNDRPYKCTVD 60
              :  :  | |:|| |::||:| | : | |||| ||: |||||  :| :| |   | |:|  :
Db         96 TSAGSTVSLSSSVESTASSTRRKTHYCDYEGCYKAFTRPSLLTEHQQTAHQGIRAYQC--E 153

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
                | : | :||||| |: |||| ||| |:|||||| :|| |:|||||||||||||  | | |
Db        154 QCGRGFTKKSHLERHLFSHSETKPFSCTVCGKGVTTRQQLRRHEITHTKSFKCPHEGCGE 213

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPA----YQCDHPGC 176
```

```
Db        214 AFYKHPQLRSHVLAVHEQKLTCTHCDKRFQRPYRLKTHIAKHHG--PASQFRYQCTNAGC 271

Qy        177 FKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVG 236

Db        272 VQCFETWSALQQHLHTDHPKLPCGVCGKLCVGETGLQMHMTVHDESRVIKNWKCSVCSDT 331

Qy        237 KFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE 296

Db        332 TYAKMADLLAHYMKTHGDSIPKELIEHVNAE-----------------SPAVVHVDEQQ 373

Qy        297 EDEEDSLDEKRSDVRSD-----SMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNN 351

Db        374 REDPEIVRPQKRRKNSDNTIMGSLQTEEKIRKLIES--GRTGLSLLLNTAGRKKRCPYLG 431

Qy        352 CDRMFSRE--YDLRRHLKWHDDNLQRIESFLNSIEKEETPEGEP 393

Db        432 CSRVFKTEDKYDL--HISKHKINDLKVKLLEDKLKTEAAAGKDP 473


RESULT 5
Q6FSJ8
ID   Q6FSJ8      PRELIMINARY;      PRT;   394 AA.
AC   Q6FSJ8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Similar to sp|P39933 Saccharomyces cerevisiae YPR186c TFC2.
GN   ORFNames=CAGL0G10021g;
OS   Candida glabrata CBS138.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=284593;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS138;
RG   Genolevures;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
DR   EMBL; CR380953; CAG59723.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 8.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
SQ   SEQUENCE   394 AA;  45716 MW;  130E6160E5FCF799 CRC64;

  Query Match           36.3%;  Score 809.5;  DB 2;  Length 394;
  Best Local Similarity   39.8%;  Pred. No. 9.7e-39;
  Matches  162;  Conservative   66;  Mismatches  134;  Indels   45;  Gaps    8;

Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRT-HSNDRPYKCTV 59

Db         13 LSRSSSVGSMSS-VASGSSVRSKTYLCEYDNCDKAFTRPSLLTEHQNTVHLGRKPWKC-- 69

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQ 119

Db         70 NQCESSFTKKIHLERHLYTHTDERPFYCSFCGKGLITRQQLKRHEVTHTKSFNCEYEGCN 129

Qy        120 EAFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGS--PAYQCDHPGCF 177

Db        130 ESFYKHPQLRAHILAVHLQSLKCHECNKSFQRPYRLKNHIAKHHNPDVVNAYQCTFSVCS 189

Qy        178 KNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK 237

Db        190 KSFKTWSALRLHVKNDHPKLKCPICSKPCVGEDGLNMHMKIHDENLVSRNWKCHICNDQS 249

Qy        238 FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEE 297

Db        250 FAKKLELLDHYSNSHSEEIPAYLLEQKVVPYVET--------NVTGKTECEK-------- 293

Qy        298 DEEDSLDEKRSDVRSDSMSAQRSIKSPTASLEGSKSVSKLISNS-GKKINCPKNNCDRMF 356

Db        294 ------SPTKKYVKATDMLAIRTETNLNRFFDGGKDAMTLLLNTVGRKFRCPYSKCYRSF 347

Qy        357 SREYDLRRHLKWHDDNLQRIESFLNS----------------IEKEE 387

Db        348 KTEEKYNIHIEKHRIHEQKLKELETSGQLQMADIEPDRQSTALEKEE 394


RESULT 6
TF3A_SCHPO
ID   TF3A_SCHPO     STANDARD;      PRT;   374 AA.
AC   Q9UTL5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Transcription factor IIIA (Factor A) (TFIIIA).
GN   Name=sfc2; ORFNames=SPAC144.09c;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=22082319; PubMed=12087160; DOI=10.1093/nar/gkf385;
RA   Schulman D.B., Setzer D.R.;
RT   "Identification and characterization of transcription factor IIIA from
RT   Schizosaccharomyces pombe.";
RL   Nucleic Acids Res. 30:2772-2781(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
```

```
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Is required for correct transcription of 5S RNA genes by
CC       RNA polymerase III. Also binds the transcribed 5S RNA's. Initiates
CC       transcription of the 5S ribosomal RNA gene.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY091590; AAM00046.1; -.
DR   EMBL; AL132675; CAB59689.1; -.
DR   PIR; T37676; T37676.
DR   HSSP; P07248; 1ARD.
DR   GeneDB_SPombe; SPAC144.09c; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 10.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 10.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW   DNA-binding; Metal-binding; Nuclear protein; Repeat; RNA-binding;
KW   Transcription regulation; Zinc-finger.
FT   ZN_FING      23     47       C2H2-type.
FT   ZN_FING      53     77       C2H2-type.
FT   ZN_FING      83    107       C2H2-type.
FT   ZN_FING     113    138       C2H2-type.
FT   ZN_FING     144    169       C2H2-type.
FT   ZN_FING     204    226       C2H2-type.
FT   ZN_FING     236    261       C2H2-type.
FT   ZN_FING     267    291       C2H2-type (atypical).
FT   ZN_FING     349    374       C2H2-type.
SQ   SEQUENCE   374 AA;  43851 MW;  7469C701FFF08FF6 CRC64;

  Query Match             22.2%;  Score 495;  DB 1;  Length 374;
  Best Local Similarity   32.8%;  Pred. No. 9.7e-21;
  Matches  123;  Conservative   58;  Mismatches  160;  Indels   34;  Gaps   13;

Qy         15 SSSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLET 74
              | :  |  | : | || | | |:||||||||||||||:||: |    | |||:||||||:
Db         13 SKNLRSAKKIFHCPYEECGKKYSRPSLLEQHLRTHSNERPFVCDYTGCSKAFYRKSHLKI 72

Qy         75 HIVSHSEKKPFHCSV--CGKGVNSRQHLKRHEITH--TKSFKCTFENCQEAFYKHQSLRH 130
                 |   |:  ||| |    |     ::|||:||   |   | : ||:| | | | ||| ||
Db         73 HKRCHTNVKPFSCHYDGCDAQFYTQQHLERHIEVHRKPKPYACTWEGCDECFSKHQQLRS 132

Qy        131 HILSVHEKTL----TCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC--FKNFQTWS 184
              || : |    |    | : |    |    || |  : |    :| | |  |   : |: ||
Db        133 HISACHTHLLPYPCTYQDCELRFATKQKLQNHVNRAHEKIISYSCPHESCVGHEGFEKWS 192

Qy        185 VLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV----GKFAK 240
               || ||:::|    |  ||:       |  |:: |   | ::     :| :      | :
Db        193 QLQNHIREAHVP-SCSICGRQFKTAAHLRHHVVLH--QTTLEERKTYHCPMEGCKKSFTR 249

Qy        241 KNELVEHYNIFHDGNIP---DDLLKETEVK-KLENLLDQGSKLNNLHELETEKLKVEEDE 296
               : | :| :: |:||:    |    :   |  |:  |::|:     |:    :  :: |
Db        250 SSALKKHISVIHEGNMAFHCDSCGTKFGYKHMLQRHLERGT-CKKAHKPYINECGIKHDG 308

Qy        297 EDEEDSLDEKRSDVRSDSMS--AQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDR 354
              :     |:|  :: |: :|  |:: |   |    | |          :: :|    |:
Db        309 IEGVAIHDQKEKELSSNLVSDVAKKIINEVTG--HGYK--------EAREYSCSFPECNY 358

Qy        355 MFSREYDLRRHLKWH 369
               | | ||: |||  |
Db        359 RFKRLYDMHRHLNSH 373


RESULT 7
```

```
Q6C2U4
ID   Q6C2U4      PRELIMINARY;      PRT;   655 AA.
AC   Q6C2U4;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Yarrowia lipolytica chromosome F of strain CLIB99 of Yarrowia
DE   lipolytica.
GN   ORFNames=YALI0F05104g;
OS   Yarrowia lipolytica CLIB99.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=284591;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   Genolevures;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382132; CAG77825.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 10.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   655 AA;  74609 MW;  5FF764F312CF4882 CRC64;

  Query Match             21.1%;  Score 470.5;  DB 2;  Length 655;
  Best Local Similarity   28.9%;  Pred. No. 4.6e-19;
  Matches  111;  Conservative   65;  Mismatches  145;  Indels   63;  Gaps   11;

Qy         21 RPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHS 80
              |   ::|  |||||||:|||||:||||:| |:| : || : | | |||:|||: |  ||:
Db         83 RTLAHVCPEEGCDKAYSRPSLLKQHLRSHYNERCFVCTYEGCGKGFFRRSHLKAHTNSHT 142

Qy         81 EKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQSLRHHILSVH---E 137
               | :||| | || |:|||| |||:||      |  ::| :       |  | :: |   :
Db        143 VAKRYHCSFCAKGFNTRQHLLRHEVTHKSIKHCPHQDCSKQCRTEVQLLEHCVADHATWK 202

Qy        138 KTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPKL 197
               |  :| :      ::  |  : |  :| |:| |  | :     |  || |: ::| :
Db        203 HQCTYPECGRAIGSERQMQYHWERDHNQTPQYKCPHDECEEVASQWIKLQKHVSKAHNRW 262

Qy        198 KCPKCGKGCVGKKGL---SSHMLSHDDST----------------MIKIWTC--DYCDVG 236
               | |       :|:      |  |: |                  : ::|||  : |
Db        263 LCHHC-------EGIFDDVQSMYEHEKSVHGLSVLYRGGHRTPGYVRQLWTCPEESCGHQ 315

Qy        237 KFAKKNELVEHYNIFHDGNIPD--DLLKETEV--------KKLENLLDQGSKLNNLHELE 286
```

```
                 |  |:   :||   :   : |  : : | ::        :| | |  :|
Db        316 SF-DKDLYAKHYVSHNKAPVRDMFESVAEYDITPEIIPYFEKPEKLTRKG---------R 365

Qy        287 TEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSG---- 342
              :|:  :|    :|| |  :|          |:| :  :     : |:    |: :|
Db        366 SERTSLEGKSGEEEYQLPSER--------PAKRQRREVSKPTNTAASIVDKITGAGYSDS 417

Qy        343 KKINCPKNNCDRMFSREYDLRRHL 366
              : | |    |   |:|:||| ||:
Db        418 RTIPCVVETCMHRFNRKYDLDRHV 441


RESULT 8
Q6R5P3
ID   Q6R5P3      PRELIMINARY;      PRT;   634 AA.
AC   Q6R5P3;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   KRAB-zinc finger protein.
GN   Name=Tce8;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129S6/SvEvTac;
RA   Brathwaite M., Waeltz P., Dudekula D., Qian Y., Nagaraja R.;
RL   Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY510701; AAR91692.1; -.
DR   HSSP; P08046; 1A1G.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 16.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 15.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 16.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.
SQ   SEQUENCE   634 AA;  73724 MW;  CBE21DDE45205D3D CRC64;

  Query Match           18.5%;  Score 411.5;  DB 2;  Length 634;
  Best Local Similarity   32.8%;  Pred. No. 1.1e-15;
  Matches  117;  Conservative   54;  Mismatches  129;  Indels   57;  Gaps   19;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  : | ||: : | |::||  |:  :|||| ::|||:|  || | ||   |: |
Db        323 KPYKC--KECGKAFTQSSSLKKHLNLHTGKKPYKC--EECDKSFTEKSTLTTHKRIHTGK 378

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLR-HHILSVHEKT 139
              ||: |::|||  ||  :|| |:  ||  | :||  : | :||    ||: |  |   ||
Db        379 KPYKCNICGKSFNSCTNLKTHQTLHTGEKPYKC--KECGKAFPYMSSLKIHQNLHPGEKP 436

Qy        140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
                ||:|:| ||  | |  |:: | |  | |:|:   | |:|   : |: |  :
Db        437 YKCKECDKSFTLKSSLRTHQIIHTGEKP-YKCNI--CGKSFNQCTNLKTHQRLHTGEKPY 493

Qy        198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIP 257
              || :|||      | ||   |     | : |  ||   | :|: |::|  | | |
Db        494 KCKECGKSFHYWSSLKSHQNLHSGE---KPYKCKECD-KSFTEKSTLIKHQRI-HTG--- 545

Qy        258 DDLLKETEVKKLE--NLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSM 315
                      |||   |: |:         |:| | |    |:
Db        546 ---------KKLYKCNICDK--SFTWCASLKTHK-KFHNGEK----------------PY 577

Qy        316 SAQRSIKSF--TASLEGSKSVSKLISNSGKKI-NCPKNNCDRMFSREYDLRRHLKWH 369
                 ::  |||   ::|: : : :   ::|:|:  |  |:||| :|     ||| | |
Db        578 KCRKCGKSFPQLSTLKRHQEIHEDKGHTGEKLYKC--NDCDRSYSYHSSFRRHQKIH 632


RESULT 9
Q6PEP4
ID   Q6PEP4      PRELIMINARY;      PRT;   634 AA.
AC   Q6PEP4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein A830058L05.
GN   Name=A830058L05;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA   Strausberg R.;
RL   Submitted (SEP-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC057947; AAH57947.1; -.
DR   HSSP; P08046; 1A1G.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 16.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 15.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 16.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.
KW   Hypothetical protein.
SQ   SEQUENCE   634 AA;  73682 MW;  E6F7E83E452DEF24 CRC64;

  Query Match           18.4%;  Score 410.5;  DB 2;  Length 634;
  Best Local Similarity   32.8%;  Pred. No. 1.2e-15;
  Matches  117;  Conservative   54;  Mismatches  129;  Indels   57;  Gaps   19;
```

```
Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        323 KPYKC--KECGKAFTQSSSLKKHLNLHTGKKPYKC--EECDKSFTEKSTLTTHKRIHTGK 378

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLR-HHILSVHEKT 139
Db        379 KPYKCNICGKSFNSCTNLKTHQTLHTGEKPYKC--KECGKAFPYMSSLKIHQNLHPGEKP 436

Qy        140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
Db        437 YKCKECDKSFTLKSSLRTHQIIHTGEKP-YKCNI--CGKSFNQCTNLKTHQRLHTGEKPY 493

Qy        198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIP 257
Db        494 KCKECGKSFHYWSSLKSHQNLHSGE---KPYKCKECD-KSFTEKSTLIKHQRI-HTG--- 545

Qy        258 DDLLKETEVKKLE--NLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSM 315
Db        546 ---------KKLYKCNICDK--SFTWCASLKTHK-KFHTGEK----------------PY 577

Qy        316 SAQRSIKSF--TASLEGSKSVSKLISNSGKKI-NCPKNNCDRMFSREYDLRRHLKWH 369
Db        578 KCRKCGKSFPQLSTLKRHQKIHEDKGHTGEKLYKC--NDCDRSYSYHSSFRRHQKIH 632


RESULT 10
Q8BI85
ID   Q8BI85      PRELIMINARY;      PRT;   634 AA.
AC   Q8BI85;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched
DE   library, clone:A830058L05 product:similar to DNA-BINDING PROTEIN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cortex;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cortex;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cortex;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cortex;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cortex;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cortex;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK043955; BAC31714.1; -.
DR   HSSP; P08045; 1ZNF.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 16.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 15.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 16.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.
KW   DNA-binding.
SQ   SEQUENCE   634 AA;  73668 MW;  AA7DEE33D87508BA CRC64;

  Query Match             18.4%;  Score 410.5;  DB 2;  Length 634;
  Best Local Similarity   32.8%;  Pred. No. 1.2e-15;
  Matches  117;  Conservative   54;  Mismatches  129;  Indels   57;  Gaps   19;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        323 KPYKC--KECGKAFTQSSSLKKHLNLHTGKKPYKC--EECDKSFTEKSTLTTHKRIHTGK 378

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLR-HHILSVHEKT 139
Db        379 KPYKCNICGKSFNSCTNLKTHQTLHTGEKPYKC--KECGKAFPYMSSLKIHQNLHPGEKP 436

Qy        140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
Db        437 YKCKECDKSFTLKSSLRTHQIIHTGEKP-YKCNI--CGKSFNQCTNLKTHQRLHTGEKPY 493

Qy        198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIP 257
Db        494 KCKECGKSFHYWSSLKSHQNLHSGE---KPYKCKECD-KSFTEKSTLIKHQRI-HTG--- 545

Qy        258 DDLLKETEVKKLE--NLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSM 315
Db        546 ---------KKLYKCNICDK--SFTWCASLKTHK-KFHTGEK----------------PY 577

Qy        316 SAQRSIKSF--TASLEGSKSVSKLISNSGKKI-NCPKNNCDRMFSREYDLRRHLKWH 369
Db        578 KCRKCGKSFPQLSTLKRHQKIHEDKGHTGEKLYKC--NDCDRSYSYHSSFRRHQKIH 632
```

```
RESULT 11
Q7RY67
ID   Q7RY67      PRELIMINARY;      PRT;   611 AA.
AC   Q7RY67;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=NCU00038.1;
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OR74A;
RA   Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA   Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA   Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA   Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA   Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA   Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA   Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA   Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA   Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA   Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,
RA   DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA   Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA   Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA   Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT   "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL   Nature 0:0-0(2003).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABX01000729; EAA27789.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 7.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
KW   Hypothetical protein.
SQ   SEQUENCE   611 AA;  67682 MW;  995B4EA5410D8932 CRC64;

  Query Match             18.3%;  Score 407;  DB 2;  Length 611;
  Best Local Similarity   24.9%;  Pred. No. 1.8e-15;
  Matches  111;  Conservative   80;  Mismatches  146;  Indels  108;  Gaps   17;

Qy          1 MSESDETKSISSLISSSSSSRP--------KKYICTYEGCDKAYNRPSLLEQHLRTHSND 52
              :: :| | ::::   :::::|         |   ||: || | :|||: |  |||:|:||
Db         91 LTHADPTTTVAATSINTTTTRASRRFPSDLKTIRCTFAGCTKTFNRPARLAAHLRSHTND 150

Qy         53 RPYKCTVDDCDKAFFRKSHLETHIV-SHSEKKPFHCSV--CGKGVNSRQHLKRHEITH-- 107
              ||:||  ||||| :  : ||  ||  ||:  : : |    |||   :   |:|| : |
Db        151 RPFKCPYDDCDKDYLEEKHLSQHIKGSHTNDRKYTCPEPGCGKSFVTNTRLRRHALVHEG 210

Qy        108 TKSFKCT-FENCQEAFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGS 166
              ::|  : :| ::| |||:|: || :||                              |
Db        211 ADRYRCRGYGDCVQSFRKHQTLQRHIRTVHL--------------------------GK 243

Qy        167 PAYQCDHPGCFKNFQTWSVLQFHIKQSHPKLK--CPKC-----GKGCVGKKG-------L 212
               || | : ||   | | | |: |::: |  ||  | :|     |    |::       |
Db        244 SAYPCGNDGCDAGFDTASALRRHVEREHGDLKFWCDECNAEGEGDDSGGRRVGFTTMLLL 303

Qy        213 SSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENL 272
              :||    ::       | :| | :   :: :: |  ::|     :|  ::|      |
Db        304 QAHMKKEHNN-------CAFCGV-RCGTQSNMMRHVELYHSAKTVED--RKTIACTWEGC 353

Qy        273 LDQGSKLNNLHELETEKLKVEEDEE---DEEDSLDEKRSDVRSDSMSAQRSIKSFTASLE 329
                : ::::||:   |      |       : |: | | |::   : : :   : ||  ::
Db        354 DKKFTRVSNLN---THIKSAHEGHRFVCGQTDTYDAKVSEIADWNFAEEGCGQGFTTRVK 410

Qy        330 GSKSVSKLISNSGKK-------------------------INCPKNNCDRMFSREYDL 362
                : |  |  : |||                       : ||   |  | |  ||
Db        411 LEEHV--LHVHLGKKRPPKLYPVPSMVAQAQQAQQALLDTRDLACPAYGCTARFIRHADL 468

Qy        363 RRHLKWHDDNLQRIESFLNSIEKEE 387
               ||::    |       | |::||
Db        469 DRHIQNDHGN--------NGIKEEE 485


RESULT 12
Q8N7Q3
ID   Q8N7Q3      PRELIMINARY;      PRT;   404 AA.
AC   Q8N7Q3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein FLJ40479.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA   Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA   Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA   Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA   Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA   Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,
RA   Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA   Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA   Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA   Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK097798; BAC05174.1; -.
DR   HSSP; P08047; 1SP2.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF00096; zf-C2H2; 14.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 14.
DR   SMART; SM00355; ZnF_C2H2; 14.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
SQ   SEQUENCE   404 AA;  45923 MW;  B1B35F62F5DF0007 CRC64;
```

```
  Query Match             17.9%;  Score 400;  DB 2;  Length 404;
  Best Local Similarity   33.9%;  Pred. No. 2.9e-15;
  Matches  121;  Conservative   39;  Mismatches  163;  Indels   34;  Gaps   19;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  | | ||:|| |:| :|   |: ::|||| ::| | |   | | ||   |:|:
Db         39 KPYKC--EECGKAFNRSSILTKHKIIHTGEKPYKC--EECGKGFSSVSTLNTHKAIHAEE 94

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSVH--EK 138
              ||: |  |||  ||   |  |:  ||  | :||  | | :||    ||  |   :|  ||
Db         95 KPYKCEECGKASNSSSKLMEHKRIHTGEKPYKC--EECGKAFSWSSSLTEH-KRIHAGEK 151

Qy        139 TLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPK 196
                |::| | | | | :||: | |  | |:|:  || | |   | |  | |
Db        152 PYKCEECGKAFNRSSILTKHKIIHTGEKP-YKCE--GCGKAFSKVSTLNTHKAIHAEEKP 208

Qy        197 LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGN 255
               || :|||       |  |   |   | | : |: |  || |:  : | ||  | | |
Db        209 YKCEECGKASNSSSKLMEHKRIH---TGEKPYKCEEC--GKAFSWSSSLTEHKRI-HAGE 262

Qy        256 IPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEE-DEEDEEDSLDEKRSDVRSDS 314
               |    |  | |         :|   :|  | :  | ||  :     |:  |   : :
Db        263 KP---YKCEECGKAFTWSSSFTKHKRIHAAE-KPYKCEECGKGFSTFSILTKHKIIHTGE 318

Qy        315 --MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                  :  |:|  |     :  |:|    |   |    | : |||   | || : |
Db        319 KRYKCEECGKAF--SWSSILTEHKIIHTGEKPYKC--EECGKAFSRSSSLTRHKRIH 371


RESULT 13
ZN43_HUMAN
ID   ZN43_HUMAN     STANDARD;      PRT;   803 AA.
AC   P17038; P28160; Q96DG1;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein
DE   KOX27).
GN   Name=ZNF43; Synonyms=KOX27, ZNF39;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=91279444; PubMed=1711675;
RA   Lovering R., Trowsdale J.;
RT   "A gene encoding 22 highly related zinc fingers is expressed in
RT   lymphoid cell lines.";
RL   Nucleic Acids Res. 19:2921-2927(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 38-190 FROM N.A.
RX   MEDLINE=91219421; PubMed=2023909;
RA   Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA   Martial J.A.;
RT   "The evolutionarily conserved Kruppel-associated box domain defines a
RT   subfamily of eukaryotic multifingered proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
RN   [4]
RP   SEQUENCE OF 476-531 FROM N.A.
RC   TISSUE=Lymphoid;
RX   MEDLINE=91145339; PubMed=2288909;
RA   Thiesen H.-J.;
RT   "Multiple genes encoding zinc finger domains are expressed in human T
RT   cells.";
RL   New Biol. 2:363-374(1990).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: T and B cell lines.
CC   -!- SIMILARITY: Belongs to the krueppel C2H2-type zinc-finger protein
CC       family.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X59244; CAA41932.1; -.
DR   EMBL; BC006528; AAH06528.2; ALT_INIT.
DR   EMBL; M61869; AAA58674.1; -.
DR   EMBL; X52358; CAA36584.1; -.
DR   PIR; S26823; S26823.
DR   HSSP; P08048; 7ZNF.
DR   TRANSFAC; T04986; -.
DR   Genew; HGNC:13109; ZNF43.
DR   MIM; 603972; -.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 21.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 22.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 22.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
KW   DNA-binding; Metal-binding; Nuclear protein; Repeat;
KW   Transcription regulation; Zinc-finger.
FT   DOMAIN       1    69       KRAB.
FT   ZN_FING    168   190       C2H2-type.
FT   ZN_FING    196   218       C2H2-type (degenerate).
FT   ZN_FING    224   246       C2H2-type (degenerate).
FT   ZN_FING    252   274       C2H2-type (degenerate).
FT   ZN_FING    280   302       C2H2-type.
FT   ZN_FING    308   330       C2H2-type.
FT   ZN_FING    336   358       C2H2-type.
FT   ZN_FING    364   386       C2H2-type.
FT   ZN_FING    392   414       C2H2-type.
FT   ZN_FING    420   442       C2H2-type.
FT   ZN_FING    448   470       C2H2-type.
FT   ZN_FING    476   498       C2H2-type.
FT   ZN_FING    504   526       C2H2-type.
FT   ZN_FING    532   554       C2H2-type.
```

```
FT   ZN_FING     560    582       C2H2-type.
FT   ZN_FING     588    610       C2H2-type.
FT   ZN_FING     616    638       C2H2-type.
FT   ZN_FING     644    666       C2H2-type.
FT   ZN_FING     672    694       C2H2-type.
FT   ZN_FING     700    722       C2H2-type.
FT   ZN_FING     728    750       C2H2-type.
FT   ZN_FING     756    778       C2H2-type.
FT   CONFLICT    181    181       P -> S (in Ref. 3).
FT   CONFLICT    712    712       P -> S (in Ref. 2).
SQ   SEQUENCE   803 AA;  93487 MW;  88C41B44DFDE2533 CRC64;

  Query Match             17.9%;  Score 400;  DB 1;  Length 803;
  Best Local Similarity   31.9%;  Pred. No. 6.3e-15;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59
                :|  |   | || ::        :   |  |  |    |  |  ||:    |  | :|   ||: ::|||
Db        367 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 422

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117
              ::| |||    |  |   |    |: :||: | ||||  |    :|   |:  ||   | :|| |
Db        423 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 480

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176
               | :|| :  :| :|  : : :|   |::| | |   ||| :||: | |  | |:|:   |
Db        481 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 537

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234
               | |  :|:|  |  |         || :|||       |::|   |    |  | : |: |
Db        538 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 593

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNLHELETEKLKVE 293
               || | : :  | |  | | |   |   |  |  |   |         :|    :|   | :  | |
Db        594 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 647

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350
              |  :  :   |     |  :   :       :|    |:|   |  :  |  :|  |:|     |   | |
Db        648 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 704

Qy        351 NCDRMFSREYDLRRHLKWH 369
                | : |:|   :|  | | |
Db        705 -CGKAFNRPSNLIEHKKIH 722


RESULT 14
Z492_HUMAN
ID   Z492_HUMAN     STANDARD;      PRT;   574 AA.
AC   Q9P255;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Zinc finger protein 492 (Fragment).
GN   Name=ZNF492; Synonyms=KIAA1473;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20277482; PubMed=10819331;
RA   Nagase T., Kikuno R., Ishikawa K.-I., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XVII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:143-150(2000).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Belongs to the krueppel C2H2-type zinc-finger protein
CC       family.
CC   -!- SIMILARITY: Contains 13 C2H2-type zinc fingers.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB040906; BAA95997.1; -.
DR   HSSP; P07248; 1ARD.
DR   Genew; HGNC:23707; ZNF492.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 12.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW   DNA-binding; Metal-binding; Nuclear protein; Repeat;
KW   Transcription regulation; Zinc-finger.
FT   NON_TER       1      1
FT   DOMAIN       15     86       KRAB.
FT   ZN_FING     184    206       C2H2-type 1.
FT   ZN_FING     212    234       C2H2-type 2.
FT   ZN_FING     240    262       C2H2-type 3.
FT   ZN_FING     268    290       C2H2-type 4.
FT   ZN_FING     296    318       C2H2-type 5.
FT   ZN_FING     324    346       C2H2-type 6.
FT   ZN_FING     352    374       C2H2-type 7.
FT   ZN_FING     380    402       C2H2-type 8.
FT   ZN_FING     408    430       C2H2-type 9.
FT   ZN_FING     436    458       C2H2-type 10.
FT   ZN_FING     464    486       C2H2-type 11.
FT   ZN_FING     492    514       C2H2-type 12 (degenerate).
FT   ZN_FING     520    542       C2H2-type 13.
SQ   SEQUENCE   574 AA;  65951 MW;  482E5F1176ECA5B0 CRC64;

  Query Match             17.9%;  Score 398;  DB 1;  Length 574;
  Best Local Similarity   30.2%;  Pred. No. 5.6e-15;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              |  |  |  |   :  | ||||  |  |  |  |  |:  :||||   ::| |||  |||  || :  |
Db        207 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 262

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : |||: |   |||  |   |  :|   |:  :||  |   |  ||  :  :|  ||:
Db        263 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 320

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
              ||    |::| |  |:: |  |  ||:  |  |    |:|:   |  |  |   |  |  |
Db        321 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 377

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
                   || :|||     |::|   |     | : |: |   |:: :  |   |  | |
Db        378 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 432

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
               |    |  |  |  ||  |  :     :|   |  :  | ||                  ::  :| :
Db        433 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 478

Qy        315 MSAQRSI-------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
              :|  :             |:|  |        :  :|  |   |  | :  |:
Db        479 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 534

Qy        362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387
              | ||  |              ||: :|  :  :   |:
Db        535 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 574
```

```
RESULT 15
O43693
ID   O43693      PRELIMINARY;      PRT;   395 AA.
AC   O43693;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Zinc-finger protein (Fragment).
GN   Name=ZFS-1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seminoma;
RX   MEDLINE=98069676; PubMed=9406578; DOI=10.1016/S0165-4608(97)00004-6;
RA   Ogawa T., Poncelet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,
RA   Udagawa K., Lecocq P., Marine J., Martial J., Hosaka M.;
RT   "Enhanced expression in seminoma of human zinc finger genes located on
RT   chromosome 19.";
RL   Cancer Genet. Cytogenet. 100:36-42(1998).
DR   EMBL; D70831; BAA24050.1; -.
DR   HSSP; P08153; 1ZFD.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 12.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   395 AA;  45291 MW;  D4E06B54DCE0BF9B CRC64;

  Query Match            17.8%;  Score 397.5;  DB 2;  Length 395;
  Best Local Similarity  35.9%;  Pred. No. 4e-15;
  Matches   97;  Conservative   35;  Mismatches   87;  Indels   51;  Gaps   14;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  : | ||:|| | |  | : |: ::|||| ::| ||  : ||| || : |: :
Db        133 KPYKC--KECGKAFNRSSTLTTHRKIHTGEKPYKC--EECGKALTQSSHLTTHKIIHTGE 188

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAF----------------- 122
              ||: |  |||  |   ||  ||: ||  | :||  | | :||
Db        189 KPYKCKKCGKAFNQSAHLTTHEVIHTGEKPYKC--EKCGKAFNHFSHLTTHKIIHTGEKP 246

Qy        123 ---------YKHQS--LRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQC 171
                      :|| |   :| |:   ||   ||:| | | : ||| :|| | | |   |:|
Db        247 YKCKECGKAFKHSSTLTKHKIIHTGEKPYKCKECGKAFNQSSKLTEHK-KIHTGEKPYEC 305

Qy        172 DHPGCFKNFQTWSVLQFHIKQSHPK---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIW 228
              :  | | |   | |  | |:|| :    || :|||       |: | : |   |  | :
Db        306 EE--CGKAFNQSSNLTRH-KKSHTEEKPYKCEECGKALNRPSTLTIHKIIH---TGEKPY 359

Qy        229 TCDYCDVGK-FAKKNELVEHYNIFHDGNIP 257
               |: |  || | : ::| :|  | | |  |
Db        360 KCEEC--GKAFNQSSKLTKHKKI-HTGEKP 386


Search completed: February  8, 2005, 15:22:03
Job time : 183 secs
```